Lfd. Nr.

# Big Data, Innovation und Datenschutz

Community Based Innovation Systems Gmbh (cbased)
SBA Research
Wirtschaftsuniversität Wien

Unter Mitarbeit von

Clemens Appl (Donau-Universität Krems)
Andreas Ekelhart (sba)
Natascha Fenz (cbased)
Peter Kieseberg (sba)
Hannes Leo (cbased, Koordination)
Sabrina Kirrane (Wirtschaftsuniversität Wien)
Axel Polleres (Wirtschaftsuniversität Wien)
Alfred Taudes (Wirtschaftsuniversität Wien)
Veronika Treitl (Wirtschaftsuniversität Wien)
Christian Singer (BMVIT)
Martin Winner (Wirtschaftsuniversität Wien)

Wien, Dezember 2017

bmvit

Bundesministerium für Verkehr, Innovation und Technologie

# Inhalt


1 Management Summary 5

2 Big Data, Innovation und Datenschutz 8

3 Datenschutz und Innovation 9

3.1 Innovation im europäischen Kontext 9

3.2 Moderne Innovationsprozesse 12

3.3 Big Data und Datenschutz 15

3.4 Die Ökonomie des Datenschutzes 16

3.5 Stylised Facts 18

3.6 Datenschutz und Innovation 20

3.6.1 Stilisierte Wirkungszusammenhänge zwischen Innovation und Datenschutz 20

3.6.2 Datenschutz, Marktstrategie und -struktur 22

4 Rechtliche Rahmenbedingungen 25

4.1 Die Datenschutz-Grundverordnung (DV-GVO) 25

4.1.1 Datenschutzrechtliche Akteure und deren Rechte und Pflichten 26

4.1.1.1 Der Betroffene - Kein Datenschutz für juristische Personen 26

4.1.2 Auftraggeber und Dienstleister nach dem DSG 2000 27

4.1.2.1 Definition und Aufgaben 27

4.1.2.2 Abgrenzung Auftraggeber – Dienstleister 28

4.1.3 Verantwortlicher und Auftragsverarbeiter nach der DS-GVO 2016/679 28

4.1.3.1 Definition und Aufgaben 28

4.1.3.2 Abgrenzung Verantwortlicher – Auftragsverarbeiter 29

4.1.4 Der datenschutzrechtliche Behördenbegriff 29

4.1.5 Personenbezogene Daten als Schutzgegenstand 30

4.1.5.1 Personenbezogene Daten nach dem DSG 2000 30

4.1.5.2 Personenbezug von Daten nach der DS-GVO 2016/679 32

4.1.6 Grundsätze der Datenverarbeitung 34

4.1.7 Zustimmung und Einwilligung 36

4.1.7.1 Zustimmung des Betroffenen 37
4.1.8 Rechtmäßige Datenverarbeitung 38
4.1.8.1 Anforderungen an die Einwilligung 38
4.1.9 Rechtliche Aspekte technischer Verarbeitungsvorgänge und der Datensicherheit 40
4.1.9.1 Bewertung rechtlicher Sonderaspekte von Big Data Anwendungen 42
4.1.9.2 Einwilligung zur Datenverarbeitung für Testzwecke 42
4.1.9.3 Opt-out 43
4.1.9.4 Einwilligung zur Datenverarbeitung im Rahmen von Big Data 44
4.1.9.5 Profiling 45
4.2 ePrivacy 45
4.2.1 Überblick über sektorspezifische Regulierungen (TKG) 45
4.2.1.1 Historische Entwicklung 45
4.2.1.2 Geltende Regelungen 46
4.2.1.3 Zentrale Fragestellungen 46
4.2.1.4 Datenverwendung nach geltendem Recht 46
4.2.1.4.1 Datenlöschung nach der Kommunikation 47
4.2.1.4.2 Verbot der Auswertung der Daten 47
4.2.1.4.3 Datenhandhabung bei Anonymisierung 47
4.2.2 Aktuelle Entwicklung auf europäischer Ebene 48
4.2.2.1 Geltende Regelungen 48
4.2.2.2 E-Privacy 48
4.2.2.3 Problematik Telekom-Unternehmen gegenüber OTT 49
4.2.2.3.1 Was sind OTT Dienste? 49
4.2.2.3.2 Worin besteht die Problematik? 49
4.2.2.4 Zeitplan 49
4.3 US vs. Europa 50
4.3.1 Bisherige Datenschutzbestimmungen im Vergleich 50
4.3.2 Situation nach Inkrafttreten der DS-GVO 52
5 Technische Analyse 55

5.1 Ausgewählte technische Aspekte des Datenschutzes 55
5.2 Löschung von Daten und Informationen 56
5.2.1 Datenlöschung in komplexen Systemen 57
5.2.2 Datenlöschung in virtualisierten Umgebungen 63
5.3 Anonymisierung von Informationen 68
5.3.1 Pseudonymisierung 70
5.3.2 Datenperturbation, Aggregation und Kataster 71
5.3.3 K-anonymity und abgeleitete Verfahren 72
5.3.4 Methoden zur Generierung von k-anonymen Datensätzen 73
5.3.5 Differential Privacy 75
5.3.6 Löschen und Anonymisierung bei Machine-Learning 75
5.3.7 Zusammenfassung 78
5.4 Privacy und Transparenz 78
5.4.1 Technische Realisierung der Transparenzschicht 81
5.4.1.1 Anforderungen an die Transparenzschicht 82
5.4.1.2 Der Status Quo 83
5.4.1.2.1 Beispiele für Local Ledger Architekturen 83
5.4.1.2.2 Beispiele für Global Ledger Architekturen mit Trusted Third Party 84
5.4.1.2.3 Beispiele für Global Ledger Architekturen mit Peer-to-Peer Netzwerk 85
5.4.1.3 Lückenanalyse im Vergleich der betrachteten Architekturen 86
5.4.1.4 Vorschläge für mögliche Lösungen 88
5.4.1.5 myData 89
5.5 Ausgewählte Methoden zur Absicherung von Informationen 90
5.5.1 Generierung synthetischer Daten 91
5.5.2 Weitergehende Anonymisierung und Löschung 91
5.5.3 Aufbau eines Research Servers 91
5.5.4 Fingerprinting zur Erkennung von Datenweitergabe 92
5.6 Forschungsfragen zu den Themen Anonymisierung und Datenlöschung 93
5.7 Einverständniserklärungen („Informed Consent“) 94

5.7.1 Anforderungen an Einverständniserklärungen 94
5.7.2 Wissenschaftlicher Status Quo 96
5.7.2.1 Methoden zur Einholung von Einverständnis 96
5.7.2.2 Anpassung von Consent 97
5.7.2.3 Einschränkungen von Einverständnis 97
5.7.3 Lückenanalyse 98
5.7.4 Mögliche Lösungsansätze 99
6 Zusammenfassung und wirtschaftspolitische Empfehlungen 101
6.1 Zusammenfassung der Erkenntnisse zur DS-GVO 103
6.1.1 Explizite Einwilligung - ex ante, ex post und darüber hinaus 103
6.1.2 Das Recht auf Vergessenwerden 105
6.1.3 Anonymisierung - Big Data ohne Einschränkungen? 107
6.1.4 Transparenz - Was ist exakt wann passiert? 108
6.2 Innovation 111
**6.3 Wirtschaftspolitische Empfehlungen** 116
7 Verwendete Literatur 124

# 1 Management Summary

Mit der Einführung der DS-GVO im Mai 2018 müssen Europäische Organisationen, die davon betroffen sind, Datenschutz und die daraus folgenden Beschränkungen beachten. Die Problematik stellt sich besonders für den Innovationstreiber Big Data, da bei derartigen Anwendungen a priori nicht klar ist, welche Daten zu sammeln sind und wie diese in entsprechende Anwendungen eingehen. In einer datenschutzfreien Welt würden derartige Anwendungen durch möglichst breite, zweckfreie Datensammlung, darauf basierende Datenanalyse zur Entdeckung verborgener Muster und die Entwicklung neuer Funktionen, die den AnwenderInnen anhand ihrer Daten angeboten werden, entstehen.

Die in vorliegender Studie vorgenommene rechtliche Analyse der DS-GVO ergibt, dass aufgrund der rechtlichen Anforderungen diese Vorgangsweise zur Entwicklung von Big Data Anwendungen nicht möglich ist: Datensammlungen und Data Mining Analysen mit personenbezogenen Daten sind ohne Einwilligung nicht erlaubt.

Dies gilt allerdings auch für Anbieter außerhalb Europas, sofern diese europäische KundInnen bedienen. Aus dieser Sicht bringt die mit Einschränkungen DS-GVO ein „level playing field“, wobei allerdings zu beachten ist, dass die dominanten US-Anbieter (Google, Facebook, Amazon etc.) im Gegensatz zu den österreichischen Klein- und Mittelbetrieben über entsprechendes Know How und Ressourcen verfügen, um die entsprechenden Anpassungen durchzuführen und bereits damit begonnen haben. Es ist daher von entscheidender Bedeutung, die Adaption durch geeignete Maßnahmen zu begleiten.

Als Ausgangspunkt zur Entwicklung eines entsprechenden Maßnahmenbündels wird in dieser Studie eine DS-GVO kompatible Vorgangsweise zur Entwicklung einer Big Data Anwendung entwickelt. Basis dieses Vorschlags sind eine rechtlichen Analyse der DS-GVO mit Schwerpunkt Big Data, eine technische Analyse der zur Umsetzung der Auflagen vorhandenen Technologien sowie Gespräche mit Unternehmen und Behörden. Grundidee der Vorgangsweise ist die Einholung der Einwilligung zur Anonymisierung und/oder Datenanalyse bereits bei der Entwicklung des datengenerierenden Systems, einer auf das Testen abgestimmten Einwilligung und einem Opt-in beim Ausrollen der Big Data Anwendung. Dadurch ist zwar ein völlig freies Datensammeln und –analysieren nicht möglich, es kann aber andererseits auch Vertrauen in den Datenschutz des Anbieters aufgebaut werden. In Verbindung mit den Informations- und Löschrechten der DS-GVO kann diese Strategie die Bereitschaft der AnwenderInnen zur Einwilligung zur Datenweitergabe fördern und bei geeigneter Umsetzung einen Vorteil gegenüber unrestringierten aber nicht transparenten Alternativen darstellen.

Entscheidend für die Praktikabilität dieses Ansatzes ist eine pragmatische Interpretation der Anforderungen an die Detailliertheit von Einwilligungserklärungen. Aufgrund der Neuheit der Materie sind diese noch unklar, die Beseitigung dieser Unsicherheiten ist daher von entscheidender Bedeutung. Damit einher geht die Schaffung von Klarheit bezüglich der rechtskonformen Regelung der übrigen Auflagen, zumal etwa das Recht auf Vergessen im Widerspruch zur Transparenz und Nachvollziehbarkeit steht. Erst auf Basis gesicherter Vorgaben kann sich die innovationsfördernde Wirkung der DS-GVO entfalten, durch die Nachfrage an neuen Technologien in den Bereichen Anonymisierung und Transparenz generiert wird. Entsprechende Technologiebereiche werden im Rahmen der Studie aufgezeigt.

Obwohl vieles ausjudiziert werden muss, gibt es auch andere Möglichkeiten mit den Unsicherheiten umzugehen. Dazu gehören weiterführende Erläuterungen oder die Option

verschiedene Ansätze mit den Behörden vorab zu diskutieren. Wenn es gelingt die Unsicherheiten abzubauen, haben Entwickler von Datenschutztechnologien mehr Anreiz in innovative Produkte und Dienstleistungen zu investieren, weil dadurch ein Markt geschaffen wird.

Dieser Ansatz ist konsistent in eine horizontale Strategie einzubauen, die die folgenden Bereiche umfasst:

- **Ausbildung**: ein Wettbewerbsvorteil durch besseren Datenschutz ist nur zu erzielen wenn die Nutzer dies auch wahrnehmen. Hierzu ist Awareness in der breiten Öffentlichkeit herzustellen und in der Ausbildung anzusetzen. Datenschutzaspekte sollten im Rahmen der Digitalisierungsstrategie „Schule 4.0“ eine wichtige Rolle spielen.
- **Forschung**: In dieser Studie wird eine Reihe von Forschungsfragen vorgestellt, durch die eine effiziente Umsetzung der DS-GVO ermöglicht wird. Diese sollten in die Grundlagenforschung und angewandte Forschung Eingang finden.
- **Förderung**: Neben der entsprechenden Forschungsförderung z.B. durch die FFG sind auch Startups, die entsprechende Lösungen anbieten, zu fördern. Hier ist insbesondere eine Verbindung zu den Blockchain Aktivitäten (https://www.blockchain-austria.gv.at/, Blockchain Village) herzustellen, zumal durch diese Technologie eine von Grund auf andere Basis für die Verarbeitung personenbezogener Daten gelegt wird.
- **Gesetzliche Rahmenbedingungen**: Wie oben aufgezeigt ist eine adäquate Auslegung der Vorschriften für die Einwilligung entscheidend für die Effizienz von Big Data Anwendungen. Eine zu detaillierte Beschreibung zieht laufende Adaptionen der Einwilligungen und Unverständnis bei den Anwendern nach sich. Im Rahmen der Möglichkeiten sollte daher auf eine pragmatische Praxis hingewirkt werden.
- Aufbau eines „MyData Local Hub“ in Österreich: Durch eine Mitgliedschaft beim MyData Projekt können mehrerer dieser Maßnahmen unterstützt werden:
    - Öffentliche Anbieter können ihre Expertise als Trusted Entitites im Rahmen eines Public Private Partnerships einbringen.
    - Den Anwendern wird bewusst welche ihrer Daten wo gespeichert sind. Dadurch wird der verantwortungsvolle Umgang mit Einwilligungen gefördert, andererseits werden die Vertrauensprobleme kleinerer Anbieter verringert.
    - Die lokale Software Community kann durch den Open Source Charakter an der Entwicklung teilhaben und wird gefördert.
    - Die Erkenntnisse können auch für die Weiterentwicklung des eGovernment verwendet werden.

Die DS-GVO ist nicht konzipiert, um Innovation zu verhindern, sondern um Datenschutz zu verbessern. Dies stellt ebenfalls eine Innovation dar, die mit beachtlichen Kosten für die Organisationen, die sie implementieren müssen, verbunden ist. Daher sollte man jetzt alle begleitenden Maßnahmen setzen, damit die Umstellung möglichst einfach über die Bühne geht und erfolgreich verläuft. Dazu gehören Informationskampagnen für Organisationen, die zur Umsetzung der DS_GVO verpflichtet sind. Aufklärung ist auch bei den KonsumentInnen notwendig, damit diese die neuen Möglichkeiten annehmen und damit umgehen können. Nur wenn informierte NutzerInnen von den neuen Möglichkeiten Gebrauch machen, kann das oft beachtete Privacy-Paradox-Privacy-Paradox (hohe Präferenz für Datenschutz bei Umfragen, aber Bereitschaft, im konkreten Fall sehr weitläufigen Datenschutzbedingungen von Anbietern zuzustimmen) verhindert werden.

Datenschutz hat aber auch eine geostrategische Dimension. Die DSGVO ermöglicht ein alternatives „business model“ für die digitale Ökonomie zu etablieren und damit einen Gegenpol zu den

marktdominanten amerikanischen Anbietern zu bilden, deren Erlösquellen sehr oft auf Datensammlung und –verwertung aufbauen. Europa ist gefordert, die strengen Datenschutzbestimmungen auch in Abkommen mit Drittstaaten einzuhalten (z.B. Privacy Shield, Fluggastdaten ). Europa muss hier auf der ganzen Linie konsequent sein, die eigenen Gesetze auf allen Ebenen einhalten und damit seine Autonomie wahren. Ebenso kann man Allianzen mit anderen Staaten außerhalb der USA, Russland und China bilden, damit diese auch die strengen europäischen Datenschutzbestimmungen einführen.

Österreich ist auf nationalstaatlicher als auch europäischer Ebene gefragt. Die konsequente und proaktive Umsetzung der DS-GVO schafft neue Geschäftsfelder. Voraussetzung ist, dass man die Umstellung möglichst reibungslos gestaltet und sowohl Unternehmen als auch KonsumentInnen informiert. Auf europäischer Ebene kann das Thema – gerade auch im Rahmen der österreichischen Präsidentschaft – weiter forciert werden.

# 2 Big Data, Innovation und Datenschutz

In einer zunehmend digitalen Gesellschaft und Wirtschaft sind der Zugang zu Daten und die damit erlaubten Handlungen ein wesentlicher Faktor, um Einsichten über die ablaufenden Prozesse zu gewinnen. Je besser diese analysiert, abgebildet und letztendlich prognostiziert werden können, desto größer ist auch der Wert der Daten für die Gestaltung von Interventionen in Wirtschaft, Politik, Verwaltung, Sicherheit, etc.

Ungehemmte Zugangs- und Verwertungsmöglichkeiten von Daten kollidieren allerdings mit gesellschaftlichen und individuellen Vorstellungen und dem Recht auf Privatsphäre. Das "Bedrohungsbild" hat sich in den vergangenen Jahren deutlich erweitert. War es vor dem Aufkommen digitaler Medien vor allem auf den Schutz vor Überwachung durch staatliche Organe und die Zugänglichmachung von privaten Information für eine breite Öffentlichkeit ausgerichtet, wurde mit der Digitalisierung über alle Gesellschafts- und Lebensbereiche die Datensammlung durch Unternehmen - insbesondere durch die US-amerikanischen Internet-Plattformen - zunehmend kritisch gesehen. Die Veröffentlichungen von Edward Snowden haben wiederum die vormals unterschätzten Möglichkeiten und Tätigkeiten der Geheimdienste aufgezeigt, und das Bedrohungsbild deutlich geschärft und erweitert.

Die Haltung zu Datenschutz und die daraus resultierenden gesetzlichen Rahmenbedingungen werden wesentlich von gesellschaftlichen Grundwerten bestimmt. Gleichzeitig haben sie auch Einfluss auf wirtschaftlichen Aktivitäten, d.h. auf den Handlungsspielraum von Unternehmen und dem Möglichkeitsraum für Innovationen. Gerade in Deutschland und Österreich sind die BürgerInnen bei der Überwachung der digitalen Kommunikation durch Unternehmen und Staaten deutlich sensibler als in anderen europäischen Ländern.

Europa hat bereits in der Vergangenheit strikte Datenschutzbestimmungen eingeführt und geht mit der neuen Datenschutz-Grundverordnung (DS-GVO) konsequent diesen Weg weiter. Die Verwendung von geschützten Daten ist nun ab Mai 2018 nur mehr mit der expliziten Zustimmung der NutzerIn, für einen bestimmten Zweck, in transparenter Weise und der Möglichkeit der Datenlöschung vorgesehen. Abgesehen von einigen sektoralen Gesetzesmaterien, hat Europa dann sehr homogene Datenschutzbestimmungen mit einem breiten Anwendungsbereich.

Der rechtliche Rahmen für Datenschutz geht in Europa deutlich weiter als in den USA. Der geringere Schutz der Privatsphäre in den USA wird häufig als Faktor angeführt, der zur Dominanz der USA im Internet und insbesondere bei online Plattformen geführt hat. Die unterschiedlichen Entwicklungsoptionen für wirtschaftliche Aktivitäten und vor allem Produkt- und Prozessinnovationen, die sich aus abweichenden datenschutzrechtlichen Bestimmungen ergeben, sind der Kern der vorliegenden Studie. Goldfarb und Tucker (2011) sehen dieses Politikfeld deshalb als Teil der Innovationspolitik, weil es nicht mehr nur darum geht missbräuchliche Verwendung von Daten zu verhindern, sondern auch die Rahmenbedingungen für die Entwicklung wirtschaftlicher Aktivitäten zu gestalten.

Insbesondere geht es dabei um Big Data. Große Datenmengen, die entweder bisher nicht analysiert oder gezielt gesammelt werden, sind der Rohstoff für die neue digitale Ökonomie - so zumindest die Ansicht von vielen ExpertInnen. Der richtige Umgang mit Big Data verspricht ein lukratives Geschäft zu werden bzw. ist es bereits. Kann man bei diesen Aussichten mit restriktiven Datenschutzbestimmungen punkten? Soll man wirtschaftliche Entwicklungschancen aufgeben, damit die BürgerInnen - die oft ohnehin selbst sehr fahrlässig mit ihren Daten umgehen - besser schlafen können? Riskiert man, dass Europa in der digitalen Plattformökonomie noch

weiter zurückfällt oder gibt es auch Möglichkeiten zu reüssieren, obwohl man strikte Datenschutzgesetze anwendet?

Diese Fragen sind erstaunlicherweise nicht mit einem Blick in wissenschaftliche Journale zu beantworten. Nicht weil es dazu wenig Forschung gäbe, sondern weil es unterschiedlichste, nicht kompatible Einsichten gibt. Die Aufgabe hier ist es also zu analysieren, über welche Wirkungskanäle Wettbewerbsvorteile bzw. -nachteile entstehen, wie groß diese sind und welche Strategien wieder zu einem “level playing field” beispielsweise zwischen den USA und Europa führen könnten.

In der Wirtschaft sind es oft die Hindernisse, auf die Unternehmen treffen, die zu neuen Lösungen - Innovationen - führen und nicht der gerade vorgezeichnete Weg, bei dem umsichtige Wirtschaftspolitiker schon alle Steine aus dem Weg geräumt haben. Wenn man vermeintliche Nachteile als Herausforderung betrachtet, entsteht Raum für den kreativen Umgang mit den Beschränkungen und für neue Lösungen. Die rechtlichen Rahmenbedingungen stecken dabei den “Streckenverlauf” ab und werden in Kapitel 4 der Studie analysiert. Berücksichtigt wird auch der Stand der Diskussion für die sektorspezifischen Datenschutzbestimmungen im Telekommunikationssektor und die Unterschiede im Regulierungsrahmen zwischen den USA und Europa.

Eine Denkschule geht davon aus, dass die rechtlichen Rahmenbedingungen die Anreize für die Entwicklung technischer Lösungen, die sowohl die Privatsphäre gewährleisten und dennoch die Analyse der Daten ermöglichen, stark erhöhen und damit Big Data-Anwendungen erlauben und neues Entwicklungspotential auftun. Vorrangig ist dabei die Entwicklung von Technologien gemeint, die nur wenige Daten benötigen, um eine gewünschte Funktionalität zur Verfügung zu stellen, die Möglichkeit Daten selektiv zu löschen und die Anonymisierung von Daten. Diese verschiedenen Optionen werden in Kapitel 55 beleuchtet.

Wie sich die veränderten rechtlichen Rahmenbedingungen und die neuen technologischen Möglichkeiten auf die Innovationsleistung auswirken können, wird in Kapitel 5 behandelt. Dabei - und das ist die wirkliche Herausforderung - sollen die Effekte einer Gesetzesmaterie die noch nicht in Kraft ist, auf Innovation untersucht werden. Ein Zusammenhang, der - wie schon angedeutet - keine klaren Muster in der einschlägigen Literatur hinterlassen hat. Um dennoch zu Aussagen zu kommen, werden die Ergebnisse der Kapitel 4 und5 mit Einsichten aus der Innovationsforschung geschnitten und so auf den Prüfstand gestellt.

Das Thema Datenschutz geht aber weit über den Innovationsaspekt und die Entwicklung von Branchen hinaus und tangiert nationales Sicherheitsstreben, militärische Strategien, politische Prozesse und gesellschaftliche Entwicklungsoptionen. Diese Zugänge und die elementaren ökonomischen Einsichten in die Kosten und Nutzen bei der (Nicht-)Weitergabe von Daten werden in Kapitel 3 knapp analysiert. Die gewonnen Einsichten aus diesem - multidisziplinären Zugang - bilden die Basis für die wirtschaftspolitischen Empfehlungen in Kapitel 6.

# 3 Datenschutz und Innovation

## 3.1 Innovation im europäischen Kontext

Die Sichtweise von Innovationsprozessen hat sich über die Jahrzehnte deutlich geändert. Die dominante Position des genialen Erfinders in den fünfziger Jahren des letzten Jahrhunderts

wurde zunehmend von den Forschungs- und Innovationsabteilungen von großen Unternehmen abgelöst. Hinzu kamen Ende der achtziger Jahre die Anwender als Innovator (van Hippel), die Vorstellung von offenen Innovationsprozessen (Chesbrough (2003)) Anfang der 2000er Jahre und zunehmend auch nicht-technische Innovationsprozesse beispielsweise in den Creative Industries aber auch Unternehmen und Sektoren. Mit der Entwicklung der Gen- und Biotechnologie (Ende der 1980er Jahre) und des Internets (Ende der 1990er Jahre) gab es eine Renaissance des Entrepreneurs, der die Möglichkeiten der neuen Technologien beleuchtet und in neu gegründeten Startups diese in Produkte und Dienstleistungen ummünzt. Letztere werden oft als Innovationsmotor gesehen, die auch großen und etablierten Unternehmen wieder neuen Innovationsgeist einhauchen können. Natürlich sind einige dieser digitalen Unternehmen mittlerweile Großkonzerne und stehen ebenfalls vor der Aufgabe, ihre Innovationsanstrengungen so zu gestalten, dass sie nicht hinter die Mitbewerber zurückfallen.

Die öffentliche Hand hat diese Innovationsprozesse auf unterschiedliche Weise unterstützt: Forschung und Lehre an Hochschulen und Universitäten bildet die Basis für viele neue (Basis-) Technologien und Personen die diese in die Industrie weitertragen können. Obwohl die Wichtigkeit von akademischer Forschung grundsätzlich etabliert war, war vieles oft weniger spektakulär als die darauf aufbauenden Innovationen im Unternehmenssektor. Mazzucatto (2013) hat hier deutlich gezeigt, wie stark die Unternehmen von öffentlich finanzierter Grundlagenforschung abhängig sind.

Ähnlich verhält es sich - auch wenn das vor allem für die USA festgehalten werden kann - für vom Militär finanzierten Projekten zum “Ausleuchten” neuer Entwicklungen, die dann aber auch für zivile Anwendungen zur Verfügung stehen. Auch hier gibt nur wenig systematische Untersuchungen die ein Gesamtbild entstehen ließen, aber dennoch einige Anhaltspunkte, die nahelegen, dass vom Militär finanzierte, ergebnisoffene Forschungs- und Entwicklungsprojekte durchaus auch zivile Effekte nach sich ziehen (Block - Keller (2011)). Die beachtliche Entwicklung des Hochtechnologiesektors in Israel war nicht zuletzt durch die Nähe zum militärischen Komplex möglich.

Europa - in Ermangelung übergeordneter (militärischer oder geopolitischer) Ziele - hat direkt auf die Förderung von Innovationen in allen Varianten und Sektoren fokussiert und dabei ein umfangreiches und diverses Förderangebot auf unterschiedlichen Governance Ebenen über eine Vielzahl von Institutionen aufgebaut. Die Förderung von Innovationen steht im Mittelpunkt der Europa 2020 Strategie der EU - ebenso wie schon in der bis 2010 gültigen Lissabon Strategie -, die letztlich darauf abzielen die Wettbewerbsfähigkeit Europas zu erhalten.

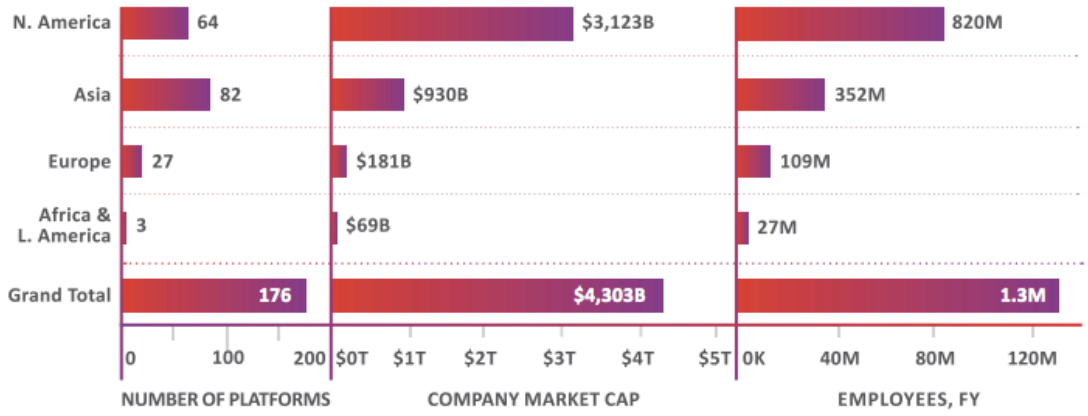
Aufgrund der großen strategischen Bedeutung von Innovationen für die Umsetzung der europäischen Strategien, sind neue Einsichten in die Natur von Innovationsprozessen immer auf fruchtbaren Boden gefallen. Um weiter wettbewerbsfähig zu bleiben, musste die neue Sichtweise in die verschiedenen Unterstützungsprogramme eingebaut werden oder gänzlich neue Programme geschaffen werden. Das hat zu einem - wie schon erwähnt - ausgesprochen ausdifferenzierten Fördersystem geführt. Neue Erkenntnisse tragen dazu bei, dass sich dieses Unterstützungssystem auch ständig an die aktuellsten Entwicklungen anpassen muss.

Die Suche nach neuen Einsichten hat auch vor Innovationsaktivitäten in Europa selbst nicht Halt gemacht. Es gibt wohl kaum eine Region, in welcher das Innovationsverhalten von Unternehmen, öffentlichen Einrichtungen etc. so intensiv untersucht wurde wie in Europa. Dennoch scheint der praktische Nutzen dieser Anstrengungen eher bescheiden. Die Erhöhung der Forschungsquote (d.h. der F&E-Ausgaben am BIP) kommt nur langsam voran und auch bei neuen technologischen Entwicklungen sind europäische Unternehmen nur selten federführend.

Im internationalen Vergleich ist erkennbar, dass Europa ist bei der Konzeption von Fertigungsverfahren und Fertigungsautomatisierung, die beispielsweise in der Automobilindustrie eingesetzt werden, weiterhin wettbewerbsfähig ist und in diesen Bereichen auch die Digitalisierung gut bewältigt hat. Europäische Unternehmen und ProgrammiererInnen haben daher vorrangig an den hier eingesetzten "Embedded Systems" gearbeitet und weniger an kommerziellen, massenmarkttauglichen Softwarepaketen.

Zu kurz kam auch die Entwicklung von Internet Plattformen. Der eklatante Rückstand zeigt sich bei der Marktkapitalisierung von digitalen Plattformen. In Asien gibt es mittlerweile 82 Plattformen mit einer Marktkapitalisierung von mehr als $1 Mrd., in den USA 64. Europa bringt es lediglich auf 27 Plattformen. Die Marktkapitalisierung ist in den USA mit $3.123 Mrd. am höchsten – ungefähr drei Mal so hoch wie jene der asiatischen Plattformen und 17-mal höher als jene der europäischen Plattformen (siehe Evans – Gawner, (2015)).

*Abbildung 1: Number, market capitalisation and employment by digital platforms*




SOURCE: Global Platform Survey, The Center for Global Enterprise, 2015

Europa hat digitale Technologien als generische Technologien verkannt und die weiteren ökonomischen und geopolitischen Implikationen unterschätzt. Wohl aus der Verbundenheit gegenüber den USA, dem ständigen Drang wettbewerbsfähig zu bleiben und dem Versuch bestehende Rückstände aufzuholen, hat Europe vor allem auf die schnelle Diffusion von digitalen Technologien gesetzt und die Appropriierung dieser vernachlässigt. Anstelle eines strategischen Ansatzes – wie etwa China oder Russland – wurde ein laissez faire Ansatz gewählt.

Die "winner takes all" Charakteristika von vielen digitalen Industrien und die Dynamik der US-amerikanischen und chinesischen Ökosysteme haben zu den bereits dargestellten europäischen Rückständen geführt, die unmittelbar kaum aufholbar scheinen. In Europa hat es deutlich länger gedauert Ökosysteme für digitale Entrepreneure aufzubauen. Auch gab es keine eigene Tradition - von Großbritannien einmal abgesehen - bei der Unterstützung von sehr riskanten Startups. Gleichzeitig war es schwierig erfolgreiche digitale Startups in Europa zu skalieren, weil der Binnenmarkt noch immer ungenügend funktioniert und damit der unmittelbar zugängliche Markt klein ist. Gleichzeitig waren europäische Startups unmittelbar und ungebremst der amerikanischen Konkurrenz ausgesetzt. Wenn trotzdem erfolgreich innoviert wurde, dann wurde das junge Unternehmen entweder von amerikanischen Finanziers aufgekauft oder hat tendenziell versucht in den USA zu skalieren.

Mittlerweile gibt es erste Anzeichen für ein Umdenken in Europa. Die Kommission – so scheint es – überdenkt ihre Digitalstrategien. Deutschland hat ein Weißbuch zum Umgang mit digitalen Plattformen erarbeitet, das vor allem auf deren Regulierung abzielt (siehe Bundesministerium für Wirtschaft und Energie (2017)). Auch die wettbewerbsrechtlichen Verfahren gegen große US-amerikanische Plattformen sind ein Anzeichen für eine neue Rolle Europas bei der Regulierung von digitalen Plattformen (siehe beispielsweise Scott (2017)).

Die strengeren europäischen Datenschutzgesetze passen gut zu den Versuchen digitale Plattformen zu regulieren, waren aber ein Vorläufer dieser Entwicklung. Dabei wird unterschätzt – unabhängig davon was die ursprünglichen Beweggründe waren –, dass die konsequente Umsetzung der neuen Datenschutz-Grundverordnung (DS-GVO) auch ein Schritt zu einem eigenständigen europäischen Ökosystem sein kann, das als Gegenstück zur totalen Überwachung durch Großkonzerne dienen kann. Die Wirkungen der DS-GVO unterscheiden sie sich deutlich von der ex post Wettbewerbsregulierung, die wenig gestalterische Elemente aufweist, sondern nur Marktmachtmissbrauch korrigieren will. Die industriepolitische Dimension der DS-GVO sollten jedenfalls als Teil einer konsequent horizontalen – d.h. alle relevanten Politikbereiche vernetzen - Digitalstrategie gesehen werden. Horizontale Strategien zeichnen sich dadurch aus, dass alle vom jeweiligen Thema tangierten Bereiche einbezogen und aufeinander abgestimmt werden. Bleibt die DS-GVO ein isolierter Baustein, weil die anderen Politikbereiche nicht andocken, dann wäre eine der wenigen Ansätze für europäischen „Landgewinn“ in der digitalen Welt vertan.

Die Datenschutzgrundverordnung ist Teil der Rahmenbedingungen für Unternehmen und bestimmt ein Stück weit was sie machen können oder eben auch nicht. Gerade in der Genese der Verordnung wurde laufend eingeworfen, dass damit Innovationen behindert werden, weil der Zugang zu Daten beschränkt wird. Tatsächlich - und das haben die bisherigen Ausführungen verdeutlicht - sind gerade die Entwicklung von digitalen Produkten und Dienstleistungen und - mit der zunehmenden Verbreitung des Lean Startup-Ansatzes - auch traditionellen Innovationen zunehmend datengetrieben. Hat sich Europa damit trotz aller erkennbaren guten Absichten ein Eigentor geschossen, da den Rückstand bei der digitalen Ökonomie noch vergrößert? Oder führen gerade die Auflagen zum Schutz der Privatsphäre dazu, dass mit Hilfe von (neu zu entwickelnden) Technologien neue wettbewerbsfähige Angebote erstellt werden, die sowohl den Datenschutz respektieren als auch von den Kunden präferiert werden? Beides ist denkbar, möglich und bereits argumentiert.

## 3.2 Moderne Innovationsprozesse

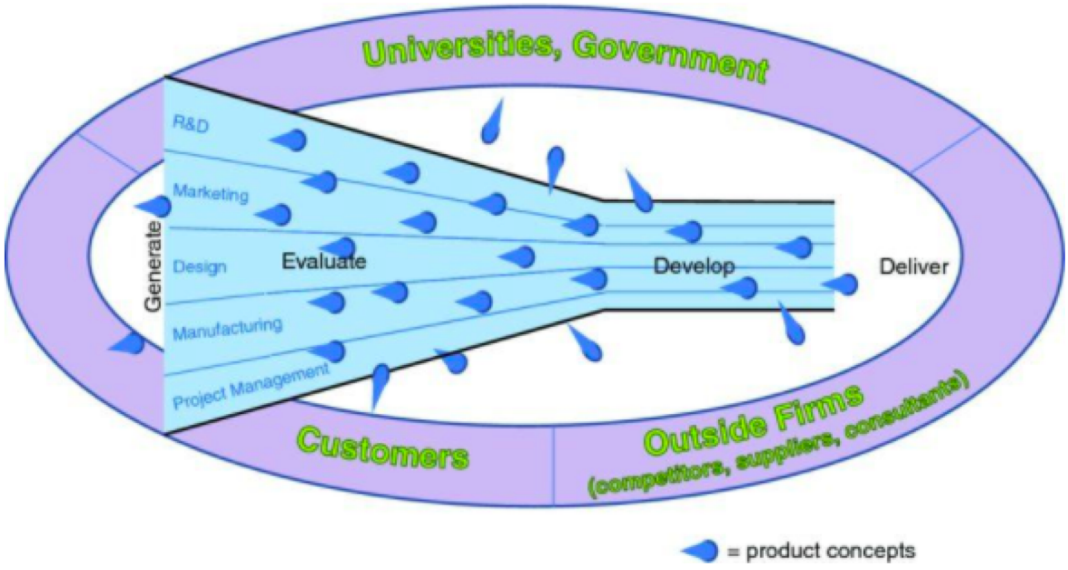
Eine fundierte Beantwortung dieser Frage ist nur bei detaillierter Betrachtung von Innovationsprozessen möglich. Innovationsprozesses werden grundsätzlich in Produkt- und Prozessinnovationen untergliedert. Diese können entweder auf industrielle Produkte oder Dienstleistungen abzielen, technologischer oder nicht technologischer Natur sein. Daneben werden auch Marketinginnovationen, organisatorische Innovationen oder neue Business Modelle als eigenständige Kategorien gesehen. Unterschieden wird auch, ob Innovationen von der Nachfrageseite – d.h. den KundInnen – oder aus den Unternehmen selbst kommen.

Innovationsprozesse sind vor allem Such- und Entdeckungsprozesse wie es Hajek (1945) zutreffend formuliert hat, bei denen imaginierte Lösungen/Innovationen in einer Art Entdeckungsverfahren auf ihre Praxistauglichkeit überprüft werden und dabei laufend überarbeitet und angepasst werden. Die Innovation, die den Such- und Entdeckungsprozessen zugrunde liegen bzw. sich aus diesen ergeben beruhen in der überwiegenden Zahl auf der

Rekombination von bestehendem Wissen (Schumpeter (1942)). Nur ein kleiner Teil technologischer Neuerungen begründet sich auf wirklich neuen Erkenntnissen. Relevantes Wissen für die Entwicklung innovativer Produkte und Dienstleistungen findet sich in unterschiedlichen Branchen, Wissensgebieten, Nutzergruppen und Stakeholdern.

Der Erfolg von Innovationen hängt wesentlich von der Kommunikation mit externen Impulsgeben und der Selektionsumwelt ab (siehe Abbildung 2). Dabei werden Ideen für neue Produkte aus unterschiedlichen Unternehmensbereichen bzw. von externen Impulsgebern in einem sogenannten Innovationstrichter selektiert. Die Beschränkungen für die erfolgreiche Entwicklung eines Produkts kommen dabei sowohl aus der Umwelt als auch aus dem Unternehmen selbst. Im Verlauf des Entwicklungsprozesses nimmt das in Entwicklung befindliche Produkt immer mehr Hürden oder die Entwicklung wird eingestellt. Letztlich "absolvieren" nur wenige der vorhandenen Ideen den Innovationsprozess erfolgreich.

*Abbildung 2:Ein offener Innovationsansatz*




*Quelle: Noble et al. (2014)*

Obwohl dieser Innovationsansatz auf den ersten Blick eine klare Vorgangsweise suggeriert, bleibt die Auswahl der Kriterien für die Selektion, und die Gruppe die diese Entscheidungen trifft, offen. Von diesen Entscheidungen wird aber das Ergebnis weitgehend determiniert. Wenn beispielsweise alle Entscheidungen vom F&E-Team getroffen werden, dann ist es wenig erstaunlich, dass das Endprodukt eher die Sichtweise dieser Gruppe widerspiegelt denn die Bedürfnisse der Kunden befriedigt.

Es kann angenommen werden, dass praktisch jedes Unternehmen Prozesse dafür entwickelt, die mit der Organisation, Kultur und Geschichte des Unternehmens zu tun haben, sodass hier die Variation groß ist. Gleichzeitig gibt es eine Unzahl an Verfahren und Techniken, um die anstehenden Informationsbedürfnisse zu befriedigen und die Entscheidungen zu stützen (Leo - Seethaler, (2017)).

Damit ist ein Kernproblem bei der Gestaltung von Innovationsprozessen angesprochen. Sind technologische Entwicklungen, Einsichten, Entdeckungen der Ausgangspunkt oder stehen die Bedürfnisse der Kunden am Anfang? Grundsätzlich spielen natürlich beide Aspekte eine Rolle

und die Herausforderung besteht darin, die technologischen Möglichkeiten auszuschöpfen ohne die Bedürfnisse der Abnehmer aus dem Auge zu verlieren.

Gerade in diesen Bereichen gibt es neue Entwicklungen:

1. Die Öffnung von Innovationsprozessen - Open Innovation und
2. Lean Startup und Behavioral Economics

Innovationsprozesse waren bis zu einem bestimmten Grad immer offen. Schon in den ersten Innovationserhebungen, die in Österreich durchgeführt wurden (1985 bzw. 1990), zeigt sich, dass unter externen Impulsgebern Konkurrenten (jeweils von mehr als >50% der Unternehmen genannt), KundInnen (>50%) sowie Messen und Kongresse (>38%) besonders wichtig waren (siehe Leo - Palme - Volk, (1992)). Die genannten Impulsgeber sind auch über die Jahre relevant geblieben. Wissenschaftliche Einrichtungen waren damals beinahe irrelevant, sind aber mittlerweile deutlich wichtiger geworden.

Das Internet erlaubt die verstärkte Einbindung von Wissensträgern und Stakeholder bei der Entwicklung von Produkten und Dienstleistungen. Prozesse, die vorher nur mit großem Aufwand und in tendenziell kleineren Communities möglich waren, können mittlerweile über breit angelegte Ausschreibungen/Calls zur Suche von Ideen oder Lösungen relativ einfach durchgeführt bzw. auf spezialisierten Plattformen aufgesetzt werden. Ebenso bietet das Internet eine Vielzahl an Suchmöglichkeiten, um Inputs für Innovationsprozesse zu generieren und Tools, um die Zusammenarbeit in Entwicklungsprozessen auf eine neue Basis zu stellen. Es ist also in Summe deutlich einfacher geworden Innovationsprozesse zu öffnen und externe Wissensträger einzubinden.

Der Lean Startup-Ansatz (siehe dazu Ries (2010), Blank - Dorf (2012), Maurya (2012)) hilft dabei Produkte und Dienstleistungen schnell zu entwickeln und dabei Kosten und Risiken zu minimieren. Dazu werden von Anfang an Hypothesen gebildet und mit unterschiedlichen Methoden (z.B. Interviews, minimum viable product (MVP)) getestet. Nur wenn der Test positiv ausfällt - es eine Validierung vom Markt gibt - wird der Prozess fortgeführt. Ein erster Meilenstein ist realisiert, wenn das „minimum viable product“ (MVP) „product market fit“ erreicht, soll heißen, dass die Nachfrage in einem hinreichend großen Marktsegment befriedigt werden kann, sodass die weiteren Entwicklungs- und Vermarktungskosten hereingespielt werden können.

Innovationsprozesse mit der Lean Startup-Methode sind deutlich schneller und effizienter als traditionelle Ansätze und funktionieren sowohl bei Startups, etablierten Unternehmen als auch in der öffentlichen Verwaltung. Lean Startup bedingt eine Öffnung des Innovationsprozesses: KundInnen, ExpertInnen und die eigenen MitarbeiterInnen sind integrale Teile des Innovationsteams. Der Ansatz ist daher mit Open Innovation-Methoden kompatibel.

Überlappungen gibt es auch zwischen Lean Startup und Behavioral Economics. Letzteres analysiert reale Entscheidungsprozesse und verlässt sich nicht auf die Verhaltens- und Rationalitätsannahmen die in ökonomischen Theorien und Modellen zum Einsatz kommen. In vielen Fällen ist menschliches Verhalten “irrational aber prognostizierbar” wie es Dan Ariely (2010) auf den Punkt gebracht hat. Die Einsichten von Behavioral Economics werden so eingesetzt, dass gewünschtes Verhalten mit weniger Reibungsverlusten erzielt wird (“Nudging”).

Diese Ansätze können und werden praktisch in allen Bereichen - von Kleinunternehmen bis hin zum öffentlichen Sektor - eingesetzt werden. Wichtig ist dabei, dass in der Regel nur dann aus einer Ideen mithilfe von Lean Startup LS und Behavioral Economics BE ein großes Unternehmen

werden kann, wenn es auch ein entsprechendes Umfeld - ein Ökosystem - gibt, das die Entrepreneure dabei unterstützt.

Startups können zumindest am Anfang nur kleine Dinge bewegen, weil ein, zwei oder drei Entrepreneure - auch wenn sie sehr clever sind - nur wenig Ressourcen haben und schnell ihre Lücke finden müssen, damit sie Risikokapital sammeln können. Bevor es diesen "product/market-fit" gibt, sind die Risikokapitalgeber eher zurückhaltend. Es braucht daher meist länger als von den meisten vermutet, bis hier ein signifikanter Beitrag erwartet werden kann.

Die Entwicklungen der amerikanischen Plattformanbieter sind hierfür gerade idealtypisch. Am Anfang geht es weniger um technologische Innovation, sondern darum ein Bedürfnis zu treffen oder ein Problem zu lösen. Die mittlerweile großen amerikanischen Plattform-Konzerne haben selten als große technologische Erneuer angefangen, sondern eine Nische besetzt und sich von dort aus verbreitert. Weder Microsoft noch Google, noch Facebook haben wirkliche technologische Sprünge gemacht, sondern Produkte entwickelt, die über die Zeit an Wettbewerbskraft gewonnen haben und von den strategischen Versäumnissen der Konkurrenz profitiert haben. Möglich waren diese Entwicklungen, weil es in den USA ein ausgebautes Ökosystem gibt, das Risikokapital für Entrepreneure zur Verfügung stellt und gleichzeitig hochqualifizierte Arbeitskräfte und Spitzenforschung vorweisen kann.

Startups sind daher zumindest in der Anfangsphase gezwungen, schnell markttaugliche Lösungen zu finden und können nicht auf ein großes Forschungsbudget zurückgreifen um neue Produkte und Dienstleistungen zu entwickeln. Natürlich - wenn alles funktioniert und man rasch wächst - dann können später fundamentale Durchbrüche realisiert werden. Startups - um es auf den Punkt zu bringen - machen in den seltensten Fällen Grundlagenforschung, sondern setzen guten Ideen mit zumeist vorhandenen Technologien möglichst schnell und kostengünstig um.

## 3.3 Big Data und Datenschutz

Wie die Erfolge von US-amerikanischen Internetfirmen wie Facebook, Google oder Amazon zeigen, ist Big Data eine aktuelle Schlüsseltechnologie, die in Zukunft weiter an Bedeutung gewinnen wird. Hierbei werden personenbezogene Daten, die im Rahmen der Interaktion mit Systemen anfallen (Suchabfragen, Klicks, Einkäufe, …) gesammelt, mittels Data Mining Verfahren auf charakteristische Patterns hin untersucht und zur Personalisierung von Angeboten (z.B. auf das Individuum zugeschnittene Werbung) verwendet. Die üblichen Grunddefinitionen von „Big Data" befassen sich hauptsächlich mit Dimensionen der Skalierbarkeit und technischen Herausforderungen immer grösser werdender Datenmengen und immer schneller zu verarbeitender Daten, um einen Wettbewerbsvorteil zu erlangen. Hier spielt es auch eine große Rolle, dass typischerweise Nicht-standard-IT Lösungen zum Einsatz kommen, um den neuen Herausforderungen gerecht zu werden. Diese Hauptherausforderungen (4 V's) anhand derer Big Data charakterisiert wird sind:

- Volume – ständig wachsende Datenmengen,
- Variety - Integration von Daten unterschiedlichster heterogener Quellen, in Datenformaten und Schemata
- Velocity – Die Geschwindigkeit der Datengenerierung und -Verarbeitung
- Value - die Schöpfung des potentiellen Mehr-Wertes durch die innovative Nutzung von Daten und das Ausreizen der Grenzen der anderen 3 V's

Im letzten „V“ - Value - liegt speziell im Umgang mit Daten mit unmittelbaren oder mittelbaren Personenbezug ein besonderer Anreiz zu Innovationen, da der Grundidee von Big Data entsprechend neue Geschäftsideen und effizientere Geschäftsprozess etwa dadurch entstehen, aus dem Datenbestand durch Data Mining neue Erkenntnisse zu gewinnen, die zur Entscheidungsfindung beim Betreiber (z.B. personalisiertes Marketing, Bonität, Betrugsverdacht etc.) oder als Basis von verbesserten Funktionen für den Anwender verwendet werden können. Im weiteren Verlauf der Verarbeitung oder durch die Verfügbarkeit von mehr Daten ergeben sich dann weitere, neue Verwendungsmöglichkeiten der Daten, wodurch ein nachhaltiger Wettbewerbsvorteil entsteht.

Genau hier ergeben sich jedoch potentielle Konflikte mit dem Datenschutz: Der datenbasierte Vergleich zwischen Datensätzen und die Einbindung und Integration bestehender und neuer Datenbestände kann es ermöglichen sensitive, personenbezogene Merkmale aus andererseits anonym erscheinenden Datenbeständen abzuleiten.

- Die BenutzerInnen haben keine Transparenz über die Art und Weise wie ihre Daten verwendet werden und an welche Stellen diese weitergegeben werden, zumal diese auf verschiedenste Anbieter verstreut sind und oft umfangreich und unscharf formulierte Einverständniserklärungen zum Einsatz kommen.
- Verpflichtungen zur Dokumentation und transparenten Aufzeichnung von personenbezogenen Daten in Echtzeit können mit hoch-optimierten und zeitkritischen Datenanalysen in Konflikt stehen.
- Verpflichtungen zum Einholen von informierter Zustimmung jeglicher Verwendung personenbezogener Daten zu einem klar definierten Verwendungszweck stehen in Konflikt neuen Datenanalysen oder der nicht zielgerichteten Entwicklung neuer, möglicherweise disruptiver Geschäftsmodelle aus bestehenden Datenbeständen („fishing in the pond“), wobei hier das konstitutive Merkmal ist, dass man im Vornherein nicht sagen kann, welche Verarbeitungsvorgänge zu einem brauchbaren Ergebnis führen.
- Know your customer (KYC), also der Anspruch, möglichst viel über den Kunden in Erfahrung zu bringen (wie es etwa im Banken- und Finanzbereich sogar eine Erfordernis durch andere EU-Regularien darstellen kann), stehen potenziell in Konflikt mit dem datenrechtlichen Grundsatz der Datenminimierung.
- Zentral gespeicherte umfangreiche Datensammlungen sind verwundbar für Hackerangriffe und Datendiebstahl.

Bei Big Data Analysen ist daher besonderes Augenmerk auf Datenschutzaspekte zu legen und eine sorgfältige Abwägung der Schutzinteressen einerseits und dem Innovationspotential andererseits zu legen. Die vorliegende Studie soll dazu einen Beitrag leisten.

## 3.4 Die Ökonomie des Datenschutzes

Datenschutz unterbindet die ungewollte Weitergabe und Verwertung von personenbezogenen Daten - die Betonung liegt auf ungewollt, weil es natürlich auch genug Gründe für eine konsensuale Datenweitergabe gibt. Es geht in diesem Kontext vor allem um die Daten von BürgerInnen. Unternehmensdaten sind von den Datenschutzbestimmungen nicht betroffen[1].

[1] Industrie 4.0 ist also von der neuen Datenschutz-Grundverordnung nur dann betroffen, wenn personenbezogene Daten zwischen Unternehmen ausgetauscht werden. Geht es beispielsweise nur um die Steuerung von Produktionsabläufen und den damitd verbundenen Daten, gibt es hier keine zusätzlichen Beschränkungen.

Unternehmen sind allerdings zumeist die Akteure, wenn es um die Weiterverarbeitung von privaten Daten geht.

Ökonomen bewerten die Weitergabe von Daten oder das Ausbleiben dieser Handlung nach den daraus entstehenden Nutzen und Kosten. Idealerweise werden dabei alle mit der Transaktion verbundenen Kosten und Nutzen und auch die indirekten Effekte oder Externalitäten - unbeabsichtigte positive und negative Nebenwirkungen - miteinbezogen. Tatsächlich ist dieses Kosten/Nutzen-Kalkül sehr oft auf einen Kernvorgang - hier die Datenweitergabe oder deren Unterlassung - beschränkt. Würden tatsächlich alle direkten und indirekten gesellschaftlichen Kosten und Nutzen miteinbezogen werden, dann wäre das grundsätzliche Entscheidungskalkül klar: wenn der Nutzen höher als die Kosten ist, dann sollte eine Aktivität erlaubt werden. Allerdings ist so eine umfassende Bewertung nur bei wenigen Problemstellungen möglich, weil selten alle Effekte miteinbezogen werden können, es schwierig ist Kosten und insbesondere Nutzen zu messen und die Betroffenheit von Wirtschaftssubjekten unterschiedlich und daher eine "one size fits all"-Lösung nicht erstrebenswert ist. Dennoch zwingt der Versuch die Kosten und Nutzen einer Transaktion zu messen, zu einer strukturierten Analyse der Problemstellung und liefert einen Beitrag zur Bewertung dieses Sachverhalts.

Wenn man diesen strikt ökonomischen Zugang auflöst, dann wird der Blick auf weitere Ebenen und Perspektiven möglich:

1. Befragungen haben ergeben dass Datenschutz bei den Betroffenen sehr positiv besetzt. Andererseits zeigt das sog. „Privacy Paradox", dass in konkreten Anwendungssituationen oft leichtfertig der Zugriff auf personenbezogenen Daten gewährt wird, indem z.B. Einverständniserklärungen pauschal und ohne gelesen zu werden akzeptiert werden.

2. Trotz des Bedürfnisses nach Datenschutz ist daher in den letzten Jahren genau das Gegenteil passiert. Die Überwachung durch Unternehmen und Staaten war noch nie so flächendeckend wie derzeit. Der Umfang an "erlaubter" Überwachung ist eine gesellschaftspolitische Fragestellung, die in der "politischen Arena" ausgehandelt werden muss. Größte Vorsicht und Zurückhaltung ist jedenfalls angesagt, weil die geringen Kosten einer flächendeckenden Überwachung sehr leicht in einen Überwachungsstaat führen bzw. der - so wie die USA das Internet überwachen - bereits Realität ist. Der Prozess dorthin ist graduell aber mehr als vorhersehbar. Nach jeder Bedrohung bzw. jedem Terroranschlag wird eine weitere Erosion von Datenschutz zumindest für den Staat verlangt, um das bereits jetzt beachtliche Arsenal von Überwachungsoptionen auszuweiten. Gleichzeitig waren die demokratischen Fundamente westlicher Demokratien schon lange nicht mehr so fragil wie derzeit. Radikale politische Umwälzungen können dazu führen, dass die neuen Machthaber alle Instrumente für die Überwachung der Bevölkerung vorfinden und davon auch hemmungslos Gebrauch machen. Generell steigt die Gefahr von Machtmissbrauch auf staatlicher Seite, wenn die neuen Instrumente einsatzbereit vorliegen.

3. Die Überwachung durch Unternehmen bzw. die Beeinflussung von demokratischen Prozessen durch Lobbys ist ebenfalls ein demokratiepolitisches Problem und verlangt mehr Aufmerksamkeit. Die sozialen Medien und die darüber gesammelten Daten, bringen weitreichende Manipulationsmöglichkeiten. Zum einen gibt es bereits glaubhafte Berichte, dass es zumindest Versuche gegeben hat die Wahl in den USA und die Brexit-Abstimmung gezielt zu manipulieren (siehe Cadwalladr (2017)). Zum anderen gibt es erste Absichten von Silicon Valley Persönlichkeiten in die Politik zu gehen. Mark

Zuckerberg hat sicherlich alle Daten um seinen Wahlkampf für das amerikanische Präsidentenamt maßzuschneidern und die Meinungsbildung auf der Plattform zu beeinflussen bzw. - wenn man dem Facebook-Manifest glaubt - die eigene Plattform für demokratische Entscheidungen zu positionieren und damit demokratische Prozesse zu privatisieren (siehe Zuckerberg, (2017)). Die grundlegende Argumentation ist schlicht: jene Meinungen, die auf Facebook hohe Zustimmung erhalten, werden üblicherweise auch von Politikern vertreten, die dann die Wahlen gewinnen. Daher kann man die Meinungsbildung gleich auf Facebook verlagern.

4. Gleichzeitig darf die geopolitische Dimension nicht aus den Augen verloren werden. Nur wer selbst in der Lage ist mit diesen Technologien kompetent umzugehen und die Entwicklung mitzubeeinflussen kann auch international mitreden. Die Reduktion auf einen Datenlieferanten - wie sie Europa in vielen Fällen eingenommen hat - ist schwer zu argumentieren. Entsprechende Sicherheits- und Militärstrategien müssen daher auch die technologische Kompetenz miteinschließen.

5. Datenschutz hat auch eine starke industriepolitische Dimension, die nicht vernachlässigt werden darf. Moderne (!) Industriepolitik entwickelt sehr stark horizontale Politikstrategien, weil die Aufgaben sich nicht an den historisch bedingten Kompetenzverteilungen zwischen den Politiksilos orientieren. Es ist daher naheliegend die Datenschutz-Grundverordnung als Teil einer europäischen Industriepolitik zu sehen, unabhängig davon, ob die Autoren das auch so gesehen haben. Moderne Industriepolitik funktioniert nur, wenn die einzelnen Elemente aufeinander abgestimmt sind. Dies ist gerade bei den Wirkungen auf Innovation von herausragender Bedeutung, weil davon abhängt, ob Märkte für neuartige Datenschutztechnologien entstehen.

Nicht akzeptabel ist daher auch, dass digitale Technologien wie jede andere generische Technologie behandelt werden. Es gibt zu vielen Externalitäten, um sich hier zurückzulehnen. Technologische Inkompetenz schränkt die Handlungsfreiheit zu stark ein.

## 3.5 Stylised Facts

Acquisti - Taylor - Wagman (2016) kommen bei Ihrer Literaturübersicht von über 250 Journalpublikationen "The Economics of Privacy" zu recht nüchternen Einsichten. Letztendlich lässt sich aus einer strikt ökonomischen Perspektive wenig Allgemeines ableiten:

1. Es kommt immer auf die spezifischen Gegebenheiten an, ob Datenschutz positive oder negative Effekte mit sich bringt. Datenschutz kann abhängig vom Kontext sowohl die private und gesamtgesellschaftliche Wohlfahrt heben oder senken.

2. Offensichtlich scheint auch, dass Personen die ihre Daten schützen bzw. die Datenweitergabe kontrollieren wollen, kaum eine rationale Entscheidung treffen können, weil sie oft gar nicht wissen, dass Daten gesammelt werden, wo diese gehandelt werden, wer diese erhält und für was sie dann eingesetzt werden.

3. Datenschutz ist auf unterschiedlichen Ebenen relevant. Ökonomen beschäftigen sich nur mit einem kleinen Ausschnitt: den Zielkonflikten, die sich aus der Weitergabe oder Zurückhaltung personenbezogener Daten ergeben können. Daneben gibt es geostrategische, militärische, sicherheits- und wirtschaftspolitische Dimensionen die wesentlich sind.

4. Es fehlt eine einheitliche und kohärente ökonomische Theorie zu Datenschutz, weil die verschiedenen Themen in unterschiedlichen Gebieten erforscht und diskutiert werden. Einige der theoretischen Einsichten sind allerdings robust.

Warum es so schwer ist, eine generelle Aussage zum Nutzen oder den Kosten von Datenschutz zu machen, ergibt sich aus den verschiedenen Anwendungsfeldern:

- Kundendaten können für Preisdiskriminierung eingesetzt werden, wenn es aufgrund der Daten möglich ist, die Zahlungsbereitschaft der KundIn abzuschätzen. Ohne Datenschutz könnten Verkäufer die Zahlungsbereitschaft - zumindest theoretisch - völlig ausreizen (Taylor, (2004); Acquisti & Varian, (2005)). Derzeit gibt es wenig Evidenz, dass individuelle Preisdifferenzierung schon im großen Stil eingesetzt wird. Uber – beispielsweise - variiert die Preise je nach Nachfrage. Die Einstieg von Amazon in den Handel geht mit Spekulationen einher, dass dort die Preise sehr stark individualisiert werden und – basierend auf den individuellen Daten über die Kundinnen – stark variieren werden. Ein Umstand der von Amazon zurückgewiesen wird (Adams (2017)).

- Kundendaten können für gezielte Werbung verwendet werden. Die NutzerInnen erhalten damit "maßgeschneiderte" Werbung. Dies sehen einige EmpfängerInnen als nützlichen Dienst, andere fühlen sich dadurch belästigt. Zielgerichtete Werbung zu Themen die eine Person interessieren, werden ausgesprochen oft als Argument gegen mehr Datenschutz angeführt. Im europäischen Kontext hat sich gezeigt, dass die schon jetzt gültige Datenschutz Grundverordnungen die Effektivität von Werbeaktivitäten deutlich senkt - konkret um 65% (Goldfarb - Tucker, (2011b)). Allerdings hat auch aufdringliche und personalisierte Werbung eine geringere Effektivität und führt zu Datenschutzbedenken der Anwender (siehe Goldfarb - Tucker, (2011a)).

- Die Weiterverwendung von Daten für Werbezwecke finanziert in vielen Bereichen Produktentwicklungen. Dazu gehören Suchmaschinen aber auch alle Gratisangebote von sozialen Netzwerken bis hin zu mobilen Apps. Die Wertschöpfung mit diesen Produktgruppen wird in einer ökonomischen Perspektive als Nutzen klassifiziert. Nicht berücksichtigt werden die Ausstrahleffekte auf andere Branchen, die dort zu Umsatzausfällen oder Beschäftigungsabbau beitragen können. Ohne wirklich umfassend Bewertung, lässt sich hier kaum abschließend schlussfolgern. Ebenso positiv wird die Wertschöpfung auf Märkten für Daten - d.h. die Aufbereitung, Aggregierung, das Zusammenspielen etc. von Daten - klassifiziert.

- Auf Kreditmärkten können mehr Informationen über die Kreditnehmer zu einem besseren funktionieren der Kreditvergabe beitragen. Ähnlich gelagert sind die Effekte bei der Personalrekrutierung, wobei gerade bei letzterer oft nicht eindeutig ist wie relevant länger zurückliegende Informationen aus sozialen Medien für die Einschätzung einer Person sind. Allerdings gibt es hier in einigen Bereichen auch Bedenken was die Weitergabe von personenbezogenen Daten betrifft. Beispielsweise geben 9 von 15 MBA-Programme die individuellen Ergebnisse der Absolventen nicht bekannt (Gottlieb und Smetters, (2011)).

Die Rolle der Datenlieferanten ist weitgehend klar auch wenn sie nicht homogen ist. Grundsätzlich kann man davon ausgehen, dass die meisten Personen Wert auf Datenschutz legen, eine Minderheit aber kein Problem hat auch personenbezogene Daten breit zu streuen oder Unternehmen zur Verfügung zu stellen.

In den USA wollten im Jahr 2009 66% der Amerikaner bzw. 86% aller jungen Erwachsene nicht, dass sie zielgerichteter Werbung ausgesetzt sind (Turow et al., (2009)). Nach dem Pew Research Center meinen 68% der Amerikaner, dass die gegenwärtigen Datenschutzbestimmungen unzureichend sind (Rainie et al., (2013)). 93% der Amerikaner geben an, dass es wichtig ist, selbst bestimmen zu können wer Zugang zu den eigenen personenbezogenen Daten hat. Nur 9% geben an, dass sie ausreichend Kontrolle über Ihre Daten haben (Madden and Rainie, (2015)). Gleichzeitig - so der Einwurf von Acquisti - Taylor - Wagman (2016) - werden Datenschutztechnologien wie TOR kaum genutzt, im Gegensatz zu datenhungrigen sozialen Netzwerken. Allerdings dürften die Datensubjekte ständig mit Zielkonflikten zwischen der Nutzung von interessanten Angeboten und der dafür notwendigen Preisgabe von personenbezogenen Informationen stehen, wobei auch Netzwerkeffekte eine Rolle spielen. Obwohl man tendenziell mehr Wert auf Datenschutz legt, (Tucker, (2012b); Stutzman et al., (2013); Kang et al., (2013); Boyd and Marwick, (2011)) gibt es Situationen in denen NutzerInnen ihre Daten nicht preisgeben und andere in denen die Zurückhaltung von Daten mit hohen Kosten verbunden ist und daher unterbleibt (Milber et al., (1995)). Die Entscheidung, Daten weiterzugeben oder nicht, ist daher in vielen Fällen kontextabhängig.

Hervorzuheben ist, dass es hier für die NutzerInnen kaum möglich ist eine rationale Entscheidung zu treffen, weil in vielen Fällen für den User unklar ist, wer Daten für welchen Zweck sammelt, weitergibt oder in anderem Kontext vermarktet. Dafür wäre ein hohes Ausmaß an technischer Kompetenz notwendig, das nur in Ausnahmefällen vorhanden ist.

Datensubjekte können von der Weitergabe von Daten profitieren: personalisierte Dienste, Nachlässe, reduzierte Suchkosten und bessere, maßgeschneiderte Informationen stehen auf der Habenseite. Vor allem aber sind es kostenlos angebotene Dienste und sozialen Netzwerke, die zur Akzeptanz von ungünstigen Nutzungsbedingungen führen.

## 3.6 Datenschutz und Innovation

Der Zusammenhang zwischen Datenschutz und Innovation wurde gerade bei der Ausarbeitung der neuen europäischen Datenschutz-Grundverordnung oft thematisiert. Den Einwänden, dass eine strikte Datenschutzgesetzgebung zu Wettbewerbsnachteilen für europäische Anbieter gerade im IKT-Bereich führt, wurde entgegengehalten, dass Nachfrager einen sicheren Umgang mit ihren Daten schätzen und sich aus diesem Umstand Wettbewerbsvorteile ergeben sollten.

Allein aus dieser - stark simplifizierten - Zusammenfassung der Diskussion kann abgeleitet werden, dass es Zusammenhänge zwischen den gesetzlichen Datenschutzregelungen und den Innovations- und Entwicklungsmöglichkeiten der Unternehmen gibt. Datenschutzregelungen sollten nach Goldfarb und Tucker (2011) daher auch als Teilgebiet der Innovationspolitik gesehen werden.

### 3.6.1 Stilisierte Wirkungszusammenhänge zwischen Innovation und Datenschutz

Der Zusammenhang zwischen Datenschutz und Innovation ist nicht endgültig geklärt, aber es gibt zumindest eine Reihe von Hypothesen (siehe dazu auch Zarsky (2015)). Die Annahmen reichen von einem schlicht positivem oder negativem Zusammenhang bis hin zu komplexeren Mustern. Aus vorhandener Evidenz können folgende Wirkungsmuster – durchaus auch gleichzeitig oder überlappend – auftreten:

1. Oft erwähnt wird die Ansicht, dass **Datenschutz Innovation ver- oder behindert**. Wenig Restriktionen beim Umgang mit sensiblen Daten erlauben es demnach Unternehmen Innovationen einzuführen, weil sie aus den Daten neue Einsichten ableiten können, und dadurch Produkte und Dienstleistungen entwickeln die exakt zu den Bedürfnissen ihrer Zielgruppe passen. Diese Vorgangsweise ist einerseits für Startups bzw. dem Lean Startup Ansatz relevant aber auch die große Phantasie bei Big Data, weil angenommen wird, dass Unternehmen zunehmend in Lage sein werden Daten zu sammeln, zu kaufen und diese auch vernünftig auszuwerten. Werden die Möglichkeiten vorhandene Daten zu nutzen durch gesetzliche Auflagen erschwert, dann ergibt sich daraus ein Innovationshemmnis.

   Diese Argumentation greift nur dann, wenn es nicht möglich ist von den NutzerInnen Einwilligungen zur Verwendung von Daten zu erhalten, weil diese in einer Weise verarbeitet werden, die nicht auf Zustimmung stößt bzw. die nachträgliche Einholung von Zustimmung an und für sich schon schwierig ist (siehe dazu Kapitel 44).

   Darüber hinaus können durch das komplexe Zusammenspiel von Datengenerierung und -übertragung positive und negative Externalitäten entstehen (z.B. frühe Warnung vor Epidemien, Überwachung, Ratings, individuelle Preisdifferenzierung, Kosten zum Schutz der privaten Daten (Datenschutztechnologien), siehe Acquisti - Taylor - Wagman (2016), S. 5)).

2. **Restriktionen bei der Weitergabe von Daten benachteiligen Business Modelle**, die ihre Innovationsaktivitäten vor allem über den **Verkauf von Daten - zumeist für Werbezwecke - finanzieren.** Wenn es Restriktionen beim Sammeln oder Weitergeben von Daten für Werbezwecke gibt, weil die Datensubjekte explizit zustimmen müssen, dann fällt dieses Business Modell zur Finanzierung von Innovationsaktivitäten ganz oder teilweise weg. Dass durch einen Wegfall der Werbeeinnahmen auch das Angebot an kostenlosen Produkten sinken wird, ist naheliegend. Damit sollten auch über Werbeeinnahmen finanzierte Innovationen zurückgehen. Diese Unternehmen haben natürlich die Möglichkeit ihr Business Modell zu ändern und ihre Produkte und Dienstleistungen zu verkaufen. Ob dieser Strategiewechsel erfolgreich sein wird, muss hier offen bleiben.

3. Es wird aber auch angenommen, dass **Datenschutz Vertrauen fördert und dadurch die Nutzung von Plattformen mit hohen Datenschutzstandards steigt**. Gleichzeitig sollten die NutzerInnen dann eher eine Nutzung ihrer personenbezogenen Daten für Innovationszwecke erlauben. Weil Vertrauen also den Zugang zu Daten erleichtert und die erhöhte Nutzung auch zu Mehreinnahmen führt, hat das auch einen positiven Effekt auf Innovationen. Unternehmen bauen also durch ihr Verhalten Reputation auf. Sobald sich bei den NutzerInnen der Eindruck gefestigt hat, dass ein Anbieter sorgfältig mit Ihren Daten umgeht, nutzen Sie dessen Dienste und ermöglichen über die gesteigerte Nachfrage Investitionen in innovative Produkte und Dienstleistungen. Gleichzeitig sind sie dann eher bereit die Nutzung ihrer Daten für Innovationsaktivitäten zuzustimmen.

   Dieses Verhaltensmuster greift schon bei der Akzeptanz von Geschäftsbedingungen. Gerade weil Geschäftsbedingungen kaum gelesen und wenn dann schwer verständlich bleiben, vertrauen NutzerInnen ihre Daten eher Unternehmen an, die ein positives Image haben oder nützliche Dienste erstellen.

Relevante Erkenntnisse für die Umsetzung einer auf besseren Datenschutz basierenden Strategie finden sich in Kelly und Murphy (2016), die auf Basis einer umfangreichen Literaturübersicht über empirische Arbeiten zur Rolle des Datenschutzes im Marketing zu folgenden Erkenntnissen kommen:

Die in Casadesus-Masanell und Hervas-Drane (2015) abgeleitete theoretische Aussage, dass durch erhöhten Datenschutz die Kundenloyalität verstärkt und die Kundeninteraktion erhöht wird findet in der Empirie Bestätigung. Der Datenschutzaspekt soll bereits ins Design des Angebots Eingang finden und auch nach außen hin authentisch dargestellt werden. Die KundInnen, denen Datenschutz ein Anliegen ist, sind auch diejenigen, die früher oder später nicht authentisches Verhalten entdecken und kommunizieren.

Vorstellbar ist natürlich auch, dass Unternehmen ihren Kunden die Wahl lassen zwischen einem werbefinanzierten Produkt, das personenbezogene Daten weitergibt und dafür kostenlos Nutzung erlaubt und einem kostenpflichtigen Produkt, bei welchem keine Daten weitergeben werden.

4. Mit Argumentationslinien 1 und 3 ist die Ansicht, dass das **Bedürfnis nach einem sicheren Umgang mit personenbezogene Daten zur verstärkten Entwicklung von Datenschutztechnologien** (Privacy Enhancing Technologies - PET) führen wird. Diese können sowohl von den Betroffenen nachgefragt werden (VPN, TOR etc.) als auch von Unternehmen, die Datenschutz gegenüber ihren Kunden garantieren wollen (siehe dazu Kapitel 4). Die gesetzlichen Rahmenbedingungen sind gerade für die Entwicklung von Datenschutztechnologien wesentlich, weil dafür Nachfrage für bestimmte Produkte und Dienstleistungen geschaffen wird.

5. Goldfarb und Tucker (2012) sehen die Auswirkungen von Datenschutz auf Innovation vor allem bei online Werbung, Gesundheitsdiensten (eHealth) und der Effizienz von internen Abläufen bei Unternehmen. Während die ersten zwei Punkte mit den bisherigen Argumentationssträngen kompatibel sind, sind die **negativen Auswirkungen auf die interne Effizienz durch die Verhinderung von Prozessinnovationen** eine neue Wirkungskette. Personenbezogene Daten werden häufig nicht nur für Produktinnovationen eingesetzt werden, sondern ermöglichen es auch interne Abläufe zu neu zu gestalten und zu optimieren. Dazu gehören Werbe- und Marketingmaßnahmen ebenso wie das Rechnungswesen. Da viele dieser Leistungen auch über Drittanbieter erstellt werden, kann es hier durchaus zu Restriktionen bei der Datenweitergabe kommen (Goldfarb - Tucker, (2012)).

### 3.6.2 Datenschutz, Marktstrategie und -struktur

Datenschutz hat massive Auswirkungen auf die möglichen Business Modelle zur Erwirtschaftung der Kosten und der darin enthaltenen Innovationsausgaben. Personenbezogene Daten, die KundInnen einem digitalen Produkt zur Verfügung stellen, ermöglichen eine höhere Servicequalität durch Individualisierung der Informationspräsentation (Fassbauer, (2014)). Weiterhin können Erträge, die durch Monetarisierung der Kundendaten auf anderen Märkten wie z.B. dem Werbemarkt erzielt werden, zur Subventionierung der Preise, die den KundInnen für die Benutzung vorgeschrieben werden, verwendet werden. Diesen positiven Effekten der Weitergabe von personenbezogenen Daten stehen der Verlust an Privatsphäre und die Reduktion der Produktqualität z.B. durch störende Werbeeinschaltungen, gegenüber.

Bei der Positionierung eines Anbieters in diesem Spannungsraum hat dieser die Einschränkungen des jeweiligen Datenschutzrechts zu beachten, wobei die technischen Möglichkeiten das Set an Gestaltungsoptionen determinieren.

Casadesus-Masanell und Hervas-Drane (2015) analysieren die Vorteilhaftigkeit von Positionierungsstrategien auf einem derartigen Markt und kommen dabei zu folgenden Ergebnissen:

Während Firmen mit Monopolmacht Erlöse sowohl aus Nutzungsentgelten und Monetarisierung von Daten gewinnen, positionieren sich im Wettbewerb die Anbieter "an den Rändern" und erzielen Erlöse entweder durch Datenmonetarisierung bei freier Produktnutzung oder durch bezahlte Angebote ohne Datenweitergabe. Je größer die Unterschiede in der Zahlungsbereitschaft desto höher die Gewinne bei beiden Positionierungen.

Je nachdem ob die durchschnittliche Zahlungsbereitschaft gering (hoch) im Verhältnis zu den Monetarisierungserlösen ist, erzielt das Unternehmen das die Daten (nicht) monetarisiert die höheren Gewinne. In Märkten mit starkem Wettbewerb und geringer Zahlungsbereitschaft der KundInnen ist daher ein hoher Grad an Datenweitergabe zu erwarten.

Bei beiden Strategien besteht die gewinnmaximierende Strategie darin, eine möglichst große Kundenbasis zu gewinnen.

Unternehmen, die keine Kundendaten monetarisieren, erhalten mehr Kundendaten.

In der Praxis ist die Lage etwas komplizierter, da der überwiegende Teil der NutzerInnen an hohen Datenschutzstandards interessiert ist. Solange keine entsprechenden gesetzlichen Datenschutzbestimmungen vorhanden sind, ist für die NutzerInnen schwer einzuschätzen, ob ein Unternehmen hohe oder niedrigen Datenschutzstandard hat. Viele NutzerInnen trauen eher großen Unternehmen. Google oder Facebook nutzen diese Tendenz durch das Schaffen von "walled gardens", ein immer umfassenderes Ökosystem, das diversifizierte Dienste anbietet und damit immer weitreichenderen Zugang zu Kundendaten hat (Kelley et al., (2010)).

Wenn Datenschutz tatsächlich funktioniert und damit zu Restriktionen bei der Auswertung von schon vorhandenen Datensätzen bei den großen Anbietern führt, dann kann Datenschutz zu geringeren Markteintrittsbarrieren und dadurch zu mehr Konkurrenz führen was wiederum zu mehr Innovation führen kann2. Bei den gegenwärtigen Marktverhältnissen – trotz der schon relativ „strengen" DS-GVO - gelingt es einigen Unternehmen eine dominante Stellung zu entwickeln und dadurch Marktmacht aufzubauen. Diese wiederum hilft dem Unternehmen noch mehr Daten zu sammeln und damit ihre Marktposition zu festigen oder noch weiter auszubauen. Google, Facebook, Amazon sind die wichtigsten Beispiele für diesen Wirkungszusammenhang.

Strikte Datenschutzbestimmungen würden dazu führen, dass diese Unternehmen ihre Datenbestände nicht mehr im vollen Ausmaß auswerten könnten, wodurch es kleineren Mitbewerbern möglich sein sollte, konkurrenzfähige Angebote zu erstellen.

Campbell et al. (2011) zeigen, dass gerade letzteres nicht der Fall sein könnte, weil große, etablierte Unternehmen eher das Vertrauen von NutzerInnen erhalten, wenn es um die Einhaltung von Datenschutzstandards geht. Die strenge Regulierung von Kreditkarten in Neuseeland ist ein empirisches Beispiel dafür: die NachfragerInnen hatten den Eindruck, dass

[2] Für diese Aussage wird angenommen, dass der Zusammenhang zwischen Wettbewerb und Innovation einen inversen U-Verlauf hat und der optimale Punkt noch nicht erreicht wurde.

nur große etablierte Betreiber die Datenschutzregelungen einhalten konnten, und haben daher neue und kleine Anbieter gemieden.

# 4 Rechtliche Rahmenbedingungen

## 4.1 Die Datenschutz-Grundverordnung (DV-GVO)

Mit der EU-Datenschutz-Grundverordnung (DS-GVO 2016/679) hat der europäische Gesetzgeber einen neuen Rechtsrahmen für den Schutz personenbezogener Daten definiert. Die DS-GVO 2016/679 trat 2016 in Kraft und ist ab Mai 2018 europaweit unmittelbar anwendbar. Das bedeutet, dass die Bestimmungen rechtlich verbindlich sind, ohne von den Mitgliedsstaaten in nationales Recht umgesetzt werden zu müssen. Nationale Usancen auf Grundlage der Datenschutz-RL (RL 95/46/EG), wie sie sich im derzeit geltenden DSG 2000 und den dazu entwickelten Grundsätzen widerspiegeln, sind damit zumindest teilweise überholt.

Ein kurzer Rückblick: Am 23. Juni 2017 endete die Begutachtungsfrist für die Regierungsvorlage zum Datenschutz-Anpassungsgesetz 2018, welches den Neuerungen durch die DS-GVO 2016/679 Rechnung tragen sollte. Das Datenschutz-Anpassungsgesetz 2018 sah umfassende Novellierungen vor: Neben der Änderung des Bundesverfassungs-Gesetzes sowie der Aufhebung des bestehenden DSG 2000 sollte im Rahmen der Öffnungsklauseln vermehrt vom nationalen Gestaltungsspielraum Gebrauch gemacht werden. Am 29. Juni 2017 wurde jedoch eine stark verkürzte Version des ursprünglichen Entwurfs vom Nationalrat beschlossen, welche später auch den Bundesrat passierte. Das somit beschlossene Gesetz hat mit dem ursprünglichen Entwurf jedoch wenig gemeinsam. Anstatt das DSG 2000 wie geplant aufzuheben wurde es (zT umfassend) geändert und in „Datenschutzgesetz (DSG)“ umbenannt. Insbesondere blieb die erwartete Änderung der Verfassungsbestimmungen aus,

Zweifelsohne zielt die DS-GVO 2016/679 auf ein hohes Datenschutzniveau ab. Wenngleich dieses ein europäisches Alleinstellungsmerkmal am internationalen Technologiemarkt sein mag, begegnen bestimmte Aspekte der DS-GVO 2016/679 durchaus auch Kritik. Bestehende oder neu entstandene Rechtsunsicherheiten sowie restriktive Regelungsansätze deuten auf ein zumindest komplexes, wenn nicht sogar schwieriges Innovationsumfeld für Technologieunternehmen hin. Obgleich die Bestimmungen der DS-GVO 2016/679 aufgrund ihres Verordnungscharakters keiner „Umsetzung“ in nationales Recht zugänglich sind, gewährt die Datenschutz-Grundverordnung durch zahlreiche Öffnungsklauseln einen zT erheblichen nationalen Regelungsspielraum. In ihren 99 Artikeln sieht die DS-GVO 2016/679 insgesamt 71 (!) Öffnungsklauseln vor, die entweder fakultativ oder obligatorisch ausgestaltet sind.

Die vorliegende Studie analysiert sowohl geltendes österreichisches Datenschutzrecht als auch den neuen Rechtsrahmen nach der DS-GVO. Relevante Fragen zur Auslegung des österreichischen Datenschutzrechts wurden bereits – zumindest teilweise – höchstgerichtlich geklärt. Mit Rechtsprechung zur DS-GVO 2016/679 kann erst mehrere Monate – wenn nicht sogar Jahre – nach dem Inkrafttreten im Mai 2018 gerechnet werden. Zahlreiche Fragen im Lichte der europäischen Verordnung eröffnen allerdings erhebliche Interpretationsspielräume, die uE nur durch einschlägige Rechtsprechung klar konturiert werden können.

## 4.1.1 Datenschutzrechtliche Akteure und deren Rechte und Pflichten

### 4.1.1.1 Der Betroffene - Kein Datenschutz für juristische Personen

**DSG 2000:** Ein wesentlicher Unterschied zwischen den Bestimmungen des DSG 2000 und jenen der DS-GVO 2016/679 stellt die Definition des Betroffenen dar. Gemäß § 4 Z 3 DSG gelten sowohl **natürliche** als auch **juristische Personen** oder Personengemeinschaften (z.B. eine Personengesellschaft) als Betroffene, wenn ihre Daten verwendet werden. Hierbei erfolgt keine Unterscheidung zwischen In- und Ausländern. Ein Datenschutzrecht für juristische Personen stellt im europäischen Vergleich eine Besonderheit dar.

Das DSG 2000 sieht zusammengefasst folgende Betroffenenrechte vor (*Jahnel*, IT-Recht[3] (2012) 445 ff):

- Informationspflicht § 24 DSG
- Auskunftsrecht § 26 DSG
- Recht auf Richtigstellung oder Löschung § 27 DSG
- Widerspruchsrecht § 28 DSG
- Verbot der automatisierten Einzelentscheidung § 49 DSG

**DS-GVO 2016/679:** Die Terminologie der DS-GVO 2016/679 weicht geringfügig dahin ab, als die Wendung „betroffene Person" angewandt wird. Davon abgesehen umfasst der persönliche Anwendungsbereich der DS-GVO 2016/679 **nur natürliche** (lebende) Personen und gewährt (nur) diesen einen Schutz der sie betreffenden Daten. Sie gewährt Schutz unabhängig von der Staatsangehörigkeit oder des Aufenthaltsortes der betroffenen Person (Plath, DSGVO[2] (2016) Art 1 Rz 3).

Eine Ausdehnung des Schutzes auf juristische Personen, wie dies durch das DSG 2000 vorgesehen war, erfolgt ausdrücklich nicht (ErwGr 14 DS-GVO). Ein Datenschutzrecht für juristische Personen stellt im europäischen Vergleich eine Besonderheit dar, da ein solches neben Österreich nämlich nur Italien und Dänemark vorsehen (*Kotschy*, (2012)). Der Ministerialentwurf des Datenschutz-Anpassungsgesetzes 2018 gibt dieses österreichische Spezifikum – weitgehend unbegründet – auf (Paal - Pauly, DS-GVO (2017) Art 4 Rz 5).

„Reine" Unternehmensdaten (zum Beispiel Firma, Adresse, Rechtsform, Jahresumsatz, Produkte, etc.) werden also nicht mehr dem Datenschutzrecht unterliegen. Nicht ausgeschlossen ist jedoch, dass die juristischen Personen betreffenden Daten auch Angaben über natürliche Personen enthalten (zum Beispiel über die UnternehmensgründerInnen, KundInnen, MitarbeiterInnen etc.). Wenn sich Daten auf identifizierbare Personen beziehen, handelt es sich wiederum um personenbezogene Daten, was zur Anwendbarkeit der DS-GVO 2016/679 führt (Paal - Pauly, DS-GVO (2017) Art 4 Rz 5 f).

Die DS-GVO 2016/679 gewährt folgende, mit dem DSG 2000 vergleichbare, Betroffenenrechte:

- Informationspflichten Art 12 – Art 14 DS-GVO
- Auskunft Art 15 DS-GVO

- Berichtigung und Löschung Art 16 – Art 17 DS-GVO
- Widerspruchsrecht Art 21 DS-GVO
- Automatisierte Entscheidung im Einzelfall Art 22 DS-GVO

Hinzu kommen folgende neue Betroffenenrechte:

- **Einschränkung der Verarbeitung Art 18 DS-GVO**

  Die betroffene Person kann vom Verantwortlichen die Einschränkung der Verarbeitung verlangen, falls (a) die Richtigkeit der personenbezogenen Daten bestritten wird, (b) die Verarbeitung unrechtmäßig ist, (c) die Daten für den Verarbeitungszweck nicht länger benötigt werden oder (d) der Betroffene Widerspruch einlegt.

- **Datenübertragbarkeit Art 20 DS-GVO**

  Die betroffene Person hat das Recht, die sie betreffenden Daten in einem maschinenlesbaren Format zu erhalten und einem anderen Verantwortlichen zu übermitteln (Jahnel, Datenschutzrecht-Update, (2016), 216).

**Selbst wenn innovative Produkte und Dienstleistungen für die Verarbeitung von Daten juristischer Personen konzipiert werden (zum Beispiel eine Software, die Konzernstrukturen und produktabhängige Umsätze grafisch aufbereitet), können diese Informationen in Ausnahmefällen auch personenbezogene Daten enthalten, die einen Rückschluss auf eine natürliche Person (zum Beispiel Name des Unternehmensgründers) erlauben. In einem solchen Fall müssten dann die datenschutzrechtlichen Bestimmungen auf diese Verarbeitung angewandt werden. Da die Eintrittswahrscheinlichkeit und die damit verbundenen Risiken für Start-ups sowie Klein- und Mittelunternehmen möglicherweise nur schwer abschätzbar sind, könnte dies ein Innovationshemmnis darstellen.**

## 4.1.2 Auftraggeber und Dienstleister nach dem DSG 2000

### 4.1.2.1 Definition und Aufgaben

Oft bedienen sich Unternehmen Dienstleistern zur Verarbeitung, Speicherung, Analyse etc. von Daten. Bei der Gestaltung der vertraglichen Beziehung zu diesen sind eine Reihe von datenschutzrechtlichen Aspekten zu beachten.

Einen **Auftraggeber** iSd § 4 Z 4 DSG treffen als „Herrn der Daten“ eine Reihe von Pflichten. Gemäß § 6 Abs 2 DSG ist der Auftraggeber für die Einhaltung der datenschutzrechtlichen Grundsätze bei jeder seiner Datenanwendungen verantwortlich, auch wenn er für die Datenanwendung einen Dienstleister heranzieht. Darüber hinaus treffen ihn die Registrierungspflicht, die Informationspflicht, die Auskunftspflicht sowie die Pflicht zur Richtigstellung und Löschung von Daten (Jahnel, IT-Recht$^3$ (2012)).

Als **Dienstleister** gilt gemäß § 4 Z 5 DSG, wer Daten nur dazu verwendet, ein in Auftrag gegebenes Werk zu erstellen. Hierbei darf der Dienstleister die Daten ausschließlich im Rahmen des Auftrags verwenden (§ 11 Abs 1 Z 1 DSG). Wenn der Auftraggeber einen Dienstleister mit der Erstellung eines Werks beauftragt und dieser Dienstleister hierzu Daten ermittelt

(sogenannter „**Ermittlungsdienstleister**“) und verwendet, bleibt er weiterhin Auftraggeber (Dohr – Pollirer – Weiss - Knyrim, DSG$^2$ (2015) § 4).

#### 4.1.2.2 Abgrenzung Auftraggeber – Dienstleister

Auftraggeber ist, wer faktisch die Entscheidung zu einer Datenverarbeitung trifft und daher nach außen als derjenige auftritt, der die Verfügungsmacht über die Daten hat. Der Auftraggeber bildet den Willensentschluss, ob und wozu eine Datenverwendung auszuführen ist. Nach dem Wortlaut des § 4 Z 5 DSG kann der beauftragte Dienstleister selbst entscheiden, inwiefern er die Daten zur Erfüllung des Werks verwendet. Der Dienstleister darf jedoch nur Detailentscheidungen selbst treffen. Diese Detailentscheidungen müssen in engem Zusammenhang mit dem aufgetragenen Werk stehen und mit dem Verarbeitungszweck in Einklang sein (Dohr – Pollirer – Weiss - Knyrim *im*, DSG$^2$ (2015) § 4).

Der Dienstleister kann vom Auftraggeber anhand der eigenverantwortlichen Entscheidungskompetenz abgegrenzt werden. Wird die **Eigenverantwortlichkeit** verneint, handelt es sich bei dem Auftragnehmer um einen datenschutzrechtlichen Dienstleister. Zu diesem Ergebnis kommt man, wenn der Auftragnehmer ausschließlich im Rahmen des Auftrags tätig wird und insbesondere den Zweck der Datenverarbeitung nicht selbst bestimmt. Liegt jedoch eigenverantwortliches Handeln vor, handelt es sich bei dem Auftragnehmer um einen datenschutzrechtlichen Auftraggeber. Eigenverantwortlichkeit ist insbesondere dann gegeben, wenn der Auftragnehmer für die überlassenen Daten ein Entgelt leistet, die Daten (auch) für einen anderen Zweck verwendet, Daten verschiedener Auftraggeber verknüpft oder über die Verwendung von Daten entgegen den Anordnungen selbst entscheidet. Dasselbe gilt, wenn der Auftragnehmer Daten trotz eines Verbots durch den Auftraggeber verwendet. Der Auftragnehmer ist in einem solchen Fall dann nicht als Dienstleister, sondern als Auftraggeber im datenschutzrechtlichen Sinn zu qualifizieren und ihn treffen die entsprechenden Pflichten des Auftraggebers (Dohr – Pollirer – Weiss - Knyrim, DSG$^2$ (2015) § 4).

### 4.1.3 Verantwortlicher und Auftragsverarbeiter nach der DS-GVO 2016/679

#### 4.1.3.1 Definition und Aufgaben

Der Begriff des Auftraggebers aus dem DSG 2000 wird in der DS-GVO 2016/679 durch jenen des **Verantwortlichen** (Art 4 Z 7 DS-GVO) ersetzt. Verantwortlicher ist jede natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet.

Durch die DS-GVO 2016/679 werden die Pflichten der Verantwortlichen stärker geregelt. Gemäß Art 24 DS-GVO muss der Verantwortliche die geeigneten technischen und organisatorischen Maßnahmen vorsehen, um eine Datenverarbeitung gemäß der DS-GVO 2016/679 sicherzustellen. Der Verantwortliche muss weiters bei der Festlegung der Mittel für die Verarbeitung und auch bei der eigentlichen Verarbeitung die geeigneten technischen und organisatorischen Maßnahmen treffen, um die Datenschutzgrundsätze (zum Beispiel Datenminimierung) umzusetzen und die Rechte der Betroffenen zu schützen (Art 25 DS-GVO). Dies umfasst insbesondere die Pflicht sicherzustellen, dass nur solche personenbezogenen Daten verarbeitet werden, deren Verarbeitung für den Zweck auch erforderlich ist.

Der **Auftragsverarbeiter** (Art 4 Z 8 DS-GVO) ersetzt den Begriff des Dienstleisters. Als Auftragsverarbeiter gilt ein Auftragnehmer, wenn er die personenbezogenen Daten im Auftrag des Verantwortlichen verarbeitet. Der Auftragsverarbeiter muss die Datenverarbeitung ausschließlich nach den Weisungen des Verantwortlichen durchführen (Art 29 DS-GVO). Er wird zumeist aufgrund eines **Vertrages** für den Verantwortlichen tätig. Gemäß Art 28 Abs 3 DS-GVO müssen Gegenstand und Dauer der Verarbeitung, Art und Zweck der Verarbeitung, die Art der personenbezogenen Daten, die Kategorien der betroffenen Personen und die Pflichten und Rechte der Verantwortlichen determiniert werden.

#### 4.1.3.2 Abgrenzung Verantwortlicher – Auftragsverarbeiter

Im Lichte der DS-GVO 2016/679 stellen sich dieselben Abgrenzungsschwierigkeiten wie schon nach dem DSG 2000. In der DS-GVO 2016/679 ist nicht näher definiert, was eine Verarbeitung **„im Auftrag“** des Verantwortlichen gemäß Art 4 Z 8 DS-GVO bedeutet. Als Abgrenzungskriterium gilt die Entscheidung über Verarbeitungszweck und Verarbeitungsmittel, die beim Verantwortlichen liegen muss. Art 28 Abs 10 DS-GVO regelt explizit, dass ein Auftragsverarbeiter als Verantwortlicher gilt, wenn er unter Verstoß gegen die DS-GVO 2016/679 die Zwecke und Mittel der Verarbeitung selbst bestimmt, eigenmächtig agiert oder gegen konkrete Weisungen oder den Vertrag verstößt. Detailentscheidungen dürfen aber an den Auftragsverarbeiter übertragen werden, ohne dass dieser zum Verantwortlichen wird. Ansonsten kann auf die obenstehenden Ausführungen zur Abgrenzung zwischen Auftraggeber und Dienstleister verwiesen werden (Knyrim, DS-GVO (2016) 170 f).

**Die Entscheidung über Art und Zweck der Datenverarbeitung obliegt dem datenschutzrechtlich Verantwortlichen, welcher der Hauptadressat der datenschutzrechtlichen Bestimmungen ist. Die korrekte rechtliche Einordnung als datenschutzrechtlich Verantwortlicher oder als Auftragsverarbeiter ist unerlässlich, damit das datenverarbeitende Unternehmen die notwendigen technischen und organisatorischen Lösungen implementieren und somit den datenschutzrechtlichen Anforderungen der Verordnung entsprechen kann. Bei Nichterfüllen der durch die DS-GVO 2016/679 auferlegten Verpflichtungen drohen empfindliche Sanktionen. Diese Abgrenzung der Verantwortungssphären hat weiters hohe Relevanz für Technologieunternehmen, weil kaum zulässige Möglichkeiten für Auftragsverarbeiter bestehen, Daten der Verantwortlichen für eigene Innovationen zu verwenden.**

### 4.1.4 Der datenschutzrechtliche Behördenbegriff

Neben Unternehmen können auch Behörden Adressaten des Datenschutzrechts sein. Die Ausführungen zum datenschutzrechtlichen Behördenbegriff gelten sowohl für das DSG 2000 als auch für die DS-GVO 2016/679.

Eine Behörde kann im Datenschutzrecht als Verantwortlicher, Auftragsverarbeiter, Empfänger oder Dritter auftreten, wobei der Begriff der Behörde weder im DSG 2000 noch in der DS-GVO 2016/679 definiert ist. Bereits nach ErwGr 31 DS-RL 95/46 obliegt es den Mitgliedstaaten festzulegen, ob eine Einrichtung, die mit der Wahrnehmung einer öffentlichen Aufgabe betraut wurde, als Behörde zu qualifizieren ist. Daran ändert die DS-GVO 2016/679 grundsätzlich nichts. Nach allgemeinem verwaltungsrechtlichen Verständnis ist der Behördenbegriff ein funktionaler. Demnach ist eine Behörde intern und/oder nach außen mit Aufgaben der öffentlichen Verwaltung betraut, welche sie durch Hoheitsakte oder Tätigkeiten der schlichten Hoheitsverwaltung erfüllt (zum Beispiel Verarbeitung personenbezogener Daten durch das Sozialamt). Beliehene

Unternehmen, also nichtöffentliche Stellen, die mit der Durchführung von Aufgaben der öffentlichen Verwaltung betraut sind, können ebenfalls Behörden sein. Der datenschutzrechtliche und der verwaltungsrechtliche Behördenbegriff sind nicht notwendigerweise deckungsgleich (Jahnel, Datenschutzrecht, (2010), 2/45 f).

Um als Behörde im datenschutzrechtlichen Sinn zu gelten, muss eine datenschutzrechtlich relevante Aufgabe gesetzlich zugewiesen werden. Auch ein organisatorisch unselbstständiger Teil einer Behörde (zum Beispiel Standesamt, Jugendamt etc.) oder eine Abteilung eines Ministeriums kann wiederum selbst eine Behörde im datenschutzrechtlichen Sinn sein. Voraussetzung ist die Zuweisung einer selbstständigen öffentlich-rechtlichen Verwaltungstätigkeit. Als Behörde wird somit eine durch Organisationsakt und nach den jeweiligen Zuständigkeitsregelungen gebildete öffentliche Stelle bezeichnet, die unabhängig vom Amtsinhaber besteht. Eine solche Stelle muss unter eigenem Namen nach außen eigenständige Aufgaben der öffentlichen Verwaltung wahrnehmen. Das bedeutet im Umkehrschluss, dass eine einzelne Verwaltungsaufgabe ihrem Träger noch keine Behördeneigenschaft verschafft. Eine unselbstständige Arbeitseinheit einer Behörde (zum Beispiel ein Referat eines Ministeriums) ist daher nicht selbst eine Behörde. Sobald die Organisationseinheit jedoch gesetzlich zugewiesene, datenschutzrechtlich relevante Aufgaben zu erfüllen hat (zum Beispiel Sozialamt), ist es eine Behörde (Gola - Schomerus, BDSG[12] (2015) § 2 Rz 6 f).

Der datenschutzrechtliche Behördenbegriff ist von großer Relevanz, weil in einer öffentlichen Stelle, abhängig von den Aufgaben im Zusammenhang mit der Datenverarbeitung, eine Vielzahl von ihnen bestehen kann. Diese können dann jeweils als Verantwortlicher, Auftragsverarbeiter, Empfänger oder Dritte in Erscheinung treten. Innerhalb einer Organisationseinheit ist in einem solchen Fall der Austausch von Daten zwischen den einzelnen datenschutzrechtlichen Behörden als eine datenschutzrechtlich relevante Übermittlung (Art 4 Z 2 DS-GVO 2016/679) an Dritte anzusehen und unterliegt daher gesetzlichen Beschränkungen.

**Eine Behörde im datenschutzrechtlichen Sinn kann sowohl als Verantwortlicher als auch als Auftragsverarbeiter tätig werden. Sie ist diesfalls Träger der datenschutzrechtlichen Rechte und Pflichten, welche die DS-GVO 2016/679 für den Verantwortlichen oder den Auftragsverarbeiter vorsieht. Innerhalb einer Organisationseinheit können mehrere datenschutzrechtliche Behörden bestehen, je nachdem, ob organisatorisch abgegrenzten Teilen (z.B. mehreren Abteilungen eines Ministeriums) jeweils eine datenschutzrechtlich relevante Aufgabe zugewiesen wird. In einem solchen Fall ist der Austausch von Daten zwischen den einzelnen Abteilungen als eine datenschutzrechtlich relevante Übermittlung anzusehen. Innovative Formen der behördlichen Datenverarbeitung müssen diesem Umstand Rechnung tragen.**

## 4.1.5 Personenbezogene Daten als Schutzgegenstand

### 4.1.5.1 Personenbezogene Daten nach dem DSG 2000

Bei Daten iSd § 4 Z 1 DSG handelt es sich um personenbezogene Angaben über Betroffene, deren Identität bestimmt oder bestimmbar ist. Personenbezogene Daten umfassen Informationen über eine natürliche Person, unabhängig davon, ob es sich um den privaten Bereich (als VerbraucherIn oder PatientIn) oder beruflichen Bereich (als MitarbeiterIn) handelt. Diese umfassen etwa Name, Geburtsdatum, Adresse, Kreditkartendaten, Kontonummer, KFZ-Kennzeichen, Geschlecht, etc. Neben Werturteilen (zum Beispiel Bonität, Charaktereigenschaften, Karriereplanung) können auch biometrische Daten (zum Beispiel

Fingerabdruck, Stimmbild, Iris) sowie Bilddaten und Tondokumente erfasst sein. Der notwendige Bezug zwischen Daten und Person ist dann gegeben, wenn sie die Identität, die Merkmale oder das Verhalten dieser Person betreffen. Selbiges gilt, wenn die Angaben dazu verwendet werden, Personen in einer bestimmten Weise zu behandeln oder zu beurteilen (Artikel-29-Datenschutzgruppe, Stellungnahme 04/2007, 10 ff).

Bei **direkt** personenbezogenen Daten ist die Identität einer Person zumindest mit hoher Wahrscheinlichkeit bestimmbar. Dies hängt von den Begleitumständen des Einzelfalls ab. Der Familienname kann zum Beispiel ausreichend sein, um den Schüler einer Klasse zu identifizieren. Im Verhältnis zur Gesamtbevölkerung eines Landes wird er aber in den meisten Fällen unzureichend sein. Es ist ausreichend, dass eine Person mit Hilfe von Zusatzinformationen bestimmt werden kann. Hierbei ist darauf abzustellen, ob eine Identifikation des/der Betroffenen durch einen **zumutbaren Ermittlungsaufwand** und durch den Einsatz von Mitteln erfolgen kann, die vernünftigerweise eingesetzt werden oder – im Gegenteil – der Mitteleinsatz ungewöhnlich wäre. Dies hängt nicht nur von der Art der Information und dem Zweck der Ermittlung (zum Beispiel Strafverfolgung) ab, sondern auch davon, ob die Technologie dem Verwender im Einzelfall ohne ungewöhnlichen Aufwand zur Verfügung steht. Insbesondere ist die dynamische **Entwicklung neuer Entschlüsselungstechnologien** zu berücksichtigen, die den Zeit- und Kostenaufwand eines Entschlüsselungsmittels signifikant senken können (Jahnel, Datenschutzrecht, (2010), 3/75 f).

Von besonderer Bedeutung ist der Begriff der **sensiblen** Daten in § 4 Z 2 DSG. Hierbei handelt es sich ausschließlich um Angaben zu natürlichen Personen, die in eine der folgenden Kategorien fallen: rassische und ethnische Herkunft, politische Meinung, religiöse oder philosophische Überzeugung, Gewerkschaftszugehörigkeit, Gesundheit und Sexualleben. Für sensible Daten gibt es zur Wahrung der schutzwürdigen Geheimhaltungsinteressen besondere Tatbestände in § 9 DSG. Außerdem hat eine Vorabkontrolle durch die Datenschutzbehörde (§ 18 Abs 2 Z 1 DSG) zu erfolgen und es können gemäß § 30 Abs 3 DSG Datenanwendungen jederzeit von der Datenschutzbehörde überprüft werden, ohne dass hierfür ein Verdacht auf Rechtswidrigkeit vorliegen muss.

Von dem Begriff der **indirekt personenbezogenen Daten** sind Informationen umfasst, die durch den/die VerwenderIn nicht oder nur mit **rechtlich unzulässigen** Mitteln auf eine Person zurückgeführt werden **können**. Gemeint ist dabei jede Form der rechtlichen Unzulässigkeit, namentlich Verstöße gegen gesetzliche oder vertragliche Pflichten. Werden Daten mittels Rechtsverstoß ermittelt, sind sie nicht als indirekt personenbezogene Daten privilegiert, sondern unterliegen in vollem Umfang dem DSG 2000. Für den/die jeweilige/n VerwenderIn von indirekt personenbezogenen Daten gelten aufgrund der geringeren Schutzwürdigkeit teilweise erleichterte datenschutzrechtliche Bestimmungen, zum Beispiel kommt es zu keiner Verletzung schutzwürdiger Gemeinhaltungsinteressen gemäß § 8 Abs 2 DSG (Bauer - Reimer, Datenschutzrecht, (2009), 123).

Pseudonymisierte Daten müssen strikt von anonymisierten Daten abgegrenzt werden. **Anonymisierte** Daten – das sind Daten, die unter keinen Umständen auf eine konkrete Person zurückgeführt werden können – unterliegen wegen mangelndem Personenbezug nicht dem DSG 2000. Dies ist etwa dann gegeben, wenn nach der Verschlüsselung die ursprünglichen Daten vollständig vernichtet werden und daher niemand mehr eine bestimmte Person identifizieren kann. Umfasst sind hierbei aber auch Fälle, bei denen die Rückführung der Daten dem Grunde nach möglich, aber wegen der Umstände des Einzelfalls äußerst unwahrscheinlich ist (in Abgrenzung zur Pseudonymisierung, wo die Re-identifikation wahrscheinlich oder gewünscht ist).

Anonymisierte Daten liegen vor, wenn die betroffene Person mit Mitteln, von deren Einsatz man vernünftigerweise ausgehen kann (folglich **ohne unverhältnismäßig hohen technischen oder wirtschaftlichen Aufwand**), nicht mehr identifiziert werden kann. Ob ein Aufwand noch verhältnismäßig ist kann in Einzelfällen anhand des erwarteten Nutzens der re-identifizierten Daten beurteilt werden. Hierbei reicht es für die Anonymität aus, dass für die Rückführbarkeit Spezialwissen erforderlich wäre, welches nur mit nicht mehr zumutbarem Aufwand erworben werden könnte (DSK 27.8.1987, 120.109, ZfVBDat 1989/7).

**Die Verarbeitung von anonymisierten Daten unterliegt mangels Personenbezug nicht dem Datenschutzrecht. Allerdings sind Daten nur dann anonymisiert, wenn sie nicht oder nur mit unvertretbarem Aufwand auf konkrete Personen zurückgeführt werden können, was die Verwertbarkeit der Informationen für innovative Datenverarbeitungskonzepte beeinträchtigen dürfte.**

**Vielfach werden in der Praxis die Vorteile der Verwendung von pseudonymisierten Daten iSv reduzierten datenschutzrechtlichen Anforderungen diskutiert. Aus juristischer Perspektive ist hierbei jedoch in mehrfacher Hinsicht Vorsicht geboten. Zum einen gibt es keine gesetzliche Definition der Pseudonymisierung im österreichischen Recht, weswegen nicht restlos geklärt ist, was unter pseudonymisierten Informationen zu verstehen ist. Es kann daher nicht abschließend beantwortet werden, ob pseudonymisierte Daten unter den Begriff der indirekt personenbezogenen Daten gemäß § 4 Z 1 DSG subsumiert werden können. Selbst wenn dies bejaht wird, bleibt fraglich, ob die datenschutzrechtlichen Erleichterungen im DSG 2000 (insbesondere die Wahrung der schutzwürdigen Geheimhaltungsinteressen, welche die Einholung einer Zustimmungserklärung obsolet machen würde) in Widerspruch zur DSRL 95/46/EG stehen und daher unangewendet bleiben müssen.**

### 4.1.5.2 Personenbezug von Daten nach der DS-GVO 2016/679

Bei personenbezogenen Daten nach Art 4 Z 1 DS-GVO handelt es sich um Informationen, die sich auf eine **identifizierte oder identifizierbare** natürliche Person („betroffene Person“) beziehen. Identifizierbar ist, wer mittels einer Kennung (zum Beispiel Vor- und Familienname) identifiziert werden kann. Zu nennen sind außerdem die Online-Kennung und Standortdaten. Durch Online-Kennungen (zum Beispiel Benutzernamen, zugewiesene IP-Adressen, etc.) kann ein bestimmter Internetnutzer identifiziert werden. Bei Standortdaten können sowohl der Aufenthaltsort (Wohnung, Arbeitsplatz, etc.) als auch Bewegungsprofile und -prognosen ermittelt werden. Die Begriffsänderung (von bestimmt/bestimmbar im DSG 2000 zu identifiziert/identifizierbar) führt uE zu keiner inhaltlichen Änderung, sondern ist lediglich auf eine verbesserte Übersetzung aus dem Englischen zurückzuführen (Knyrim, DS-GVO, (2016)).

Durch die DS-GVO 2016/679 kam es im Bereich der Datenkategorien im Wesentlichen zu terminologischen und nur teilweise zu inhaltlichen Änderungen. Art 9 DS-GVO regelt die Verarbeitung **besonderer Kategorien personenbezogener Daten**. Die Verordnung nennt hier in Ergänzung zu den „sensiblen Daten“ gemäß § 4 Z 2 DSG auch genetische und biometrische Daten (ErwGr 10 DS-GVO nennt „sensible Daten“ als Synonym für die besonderen Kategorien personenbezogener Daten). Die Verarbeitung dieser Informationen unterliegt auch nach der DS-GVO 2016/679 strengeren datenschutzrechtlichen Bestimmungen, wie etwa dem Erfordernis einer ausdrücklichen Einwilligung gem. Art 9 Abs 2 Z 1 DS-GVO. Das Gesetz gibt allerdings keine Anleitung, wie mit **potenziell** sensiblen Daten umzugehen ist. Hierbei handelt es sich um Angaben, bei denen der Auftraggeber erst im Zuge der Datenauswertung erkennt, ob besondere

Kategorien personenbezogener Daten enthalten sind. Dies ergibt sich etwa bei der Auswertung von Logfiles über die Verwendung des WWW.

Die DS-GVO 2016/679 kennt, im Gegensatz zum DSG 2000, keine Sonderform der indirekt personenbezogenen Daten. Obgleich **pseudonymisierte Daten** in Art 4 Z 5 DS-GVO legal definiert werden, sieht die Verordnung auch für sie keine erleichterten Bestimmungen vor. Unter Pseudonymisierung ist die Verarbeitung personenbezogener Daten zu verstehen, sodass die Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen Person zugeordnet werden können. Als personenbezogene Daten unterliegen sie in vollem Umfang der DS-GVO 2016/679. Bedeutung können pseudonymisierte Daten allerdings bei der Beurteilung der Rechtmäßigkeit der Datenverarbeitung (zum Beispiel die Zulässigkeit einer Zweckänderung (Art 6 Abs 4 lit e DS-GVO) erlangen.

In der DS-GVO 2016/679 findet sich keine mit dem österreichischen Datenschutzrecht vergleichbare Privilegierung für die Verarbeitung von pseudonymisierten Daten. Daher ist die Diskussion, ob es zu einem Wegfall der Notwendigkeit einer Zustimmungserklärung bei der Verarbeitung von pseudonymisierten Daten kommt, für den Geltungsbereich der DS-GVO 2016/679 obsolet.

**Anonymisierung** bedeutet, dass personenbezogene Daten verändert werden, sodass die Identifikation einer Person weder direkt, indirekt noch unter Zuhilfe-nahme zusätzlicher Informationen möglich ist. Eine scharfe Trennlinie zwischen Anonymisierung und Pseudonymisierung kann auch im Lichte der DS-GVO nicht gezogen werden. Der Unterschied liegt darin, dass bei einer Pseudonymisierung eine spätere **Re-Identifikation** der Daten durchaus erwünscht sein kann (und somit auch möglich sein muss), während bei der Anonymisierung ebendiese Rückführbarkeit mit hoher Wahrscheinlichkeit ausgeschlossen werden soll. Der Begriff der „**hohen Wahrscheinlichkeit**“, mit welcher die Rückführbarkeit der Daten ausgeschlossen werden muss, ist Gegenstand des juristischen Diskurses. Hierbei sind nicht nur die allgemeine Lebenserfahrung sowie das vorhandene oder erwerbbare Zusatzwissen des Datenverwenders zu berücksichtigen, sondern auch aktuelle und künftige technische Möglichkeiten der Datenverarbeitung sowie Zeit- und Kostenaufwand (*Laue/Nink/Kremer*, Das neue Datenschutzrecht in der betrieblichen Praxis, (2016) § 1 Rz 21 ff).

Hervorzuheben ist, dass das Anonymisieren von Daten bereits eine Datenverarbeitung (Art 4 Z 2 DS-GVO) darstellt, weswegen dieser Vorgang den datenschutzrechtlichen Bestimmungen unterliegt. Dies bedeutet, dass die Anonymisierung nur mit Zustimmung erfolgen darf oder ein anderer Erlaubnistatbestand des Art 6 Abs 1 DS-GVO erfüllt sein muss. Werden jedoch bereits anonymisierte Daten erhoben, unterliegt diese Erhebung nicht der DS-GVO (Knyrim, DS-GVO (2016), 48).

Insbesondere Methoden der Anonymisierung personenbezogener Daten können raschen technologischen Entwicklungen unterliegen. Technische Standards, die heute noch die Anforderungen an eine sichere Anonymisierung erfüllen, können aufgrund der Entwicklung innovativer Entschlüsselungsmethoden schon in wenigen Monaten obsolet sein. Die Darstellung des rechtlichen Rahmens als Grundlage für die Entwicklung gesetzeskonformer technischer Lösungen der Datenverarbeitung muss daher stets zukunftsorientiert erfolgen. Unternehmen sind jetzt gefragt, **state-of-the-art Technologien** unter Berücksichtigung der gesetzlichen Restriktionen zu implementieren, um die Bestimmungen der DS-GVO 2016/679 ab ihrem Inkrafttreten 2018 zu erfüllen.

**Der Gesetzgeber lässt technische Aspekte im Bereich Datenschutz bewusst offen und beschränkt sich auf die Festlegung allgemeiner Grundsätze, um zukünftigen technologischen Entwicklungen Rechnung zu tragen. Die Auswahl geeigneter Anonymisierungsverfahren – insbesondere in Hinblick darauf, welches Re-Identifikationsrisiko in Kauf genommen wird – ist somit immer auch eine unternehmerische Entscheidung und stellt Teil des unternehmerischen Risikos dar. Innovative Unternehmen müssen daher regelmäßig überprüfen, ob es maßgebliche technologische Neuerungen gibt, welche die Rückführung ehemals anonymer Daten ermöglicht. Ist das der Fall, ist ab diesem Zeitpunkt die DS-GVO 2016/679 für die diese Daten betreffenden Verarbeitungen anwendbar.**

**Technologische Maßnahmen sollen daher immer auch darauf abzielen, die nachträgliche Re-Identifikation (möglichst) auszuschließen. Insbesondere kann sich eine nachträgliche Bestimmbarkeit durch einen technologischen Fortschritt ergeben, wenn dadurch Zeit und Kosten der Re-Identifikation signifikant gesenkt werden (ErwGr 26). In einer zeitlichen Betrachtung ist daher festzustellen, dass heute viel seltener von anonymisierten Daten gesprochen werden kann als noch vor einigen Jahren, da die technischen Möglichkeiten, scheinbar anonyme Daten einer bestimmten Person zuzuordnen, stark gewachsen sind. Problematisch wird dies insbesondere für Unternehmen, die innovative Lösungskonzepte für die Verarbeitung anonymer Daten entwickeln (und somit nicht in den Anwendungsbereich der DS-GVO 2016/679 fallen), die Wiederherstellung des Personenbezugs durch technische Neuerungen jedoch zu einem späteren Zeitpunkt wieder möglich ist. Diese Unternehmen sehen sich dann mit den datenschutzrechtlichen Regelungen konfrontiert, ohne dies zuvor antizipieren zu können.**

**Für Technologieunternehmen stellt sich im Zusammenhang mit Big Data Analysen mitunter das Problem, dass aus nicht-sensiblen Daten durch Anwendung bestimmter Algorithmen sensible Daten gewonnen werden können. Insofern können im Pool nicht-sensibler Daten zugleich sensible Daten bestehen, die zwar noch nicht manifest aber bereits mittelbar im Datenpool angelegt sind. Damit können unter Umständen Restriktionen bezüglich sensibler Daten bereits auf gepoolte nicht-sensible Daten durchschlagen, was einen nicht unerheblichen Einfluss auf den Innovationsprozess nahelegt. In einem Datenpool können Informationen zum Einkaufsverhalten, Bewegungsprofile sowie Internetsuchergebnisse enthalten sein, welche idR in die Kategorie der nicht-sensiblen Daten fallen. Werden diese Informationen einer natürlichen Person zugeordnet, lassen sich mitunter Aussagen zu deren Gesundheitszustand treffen (z.B. Vorliegen einer Schwangerschaft, Diagnose einer Krankheit), was aus datenschutzrechtlicher Perspektive eine sensible Information darstellt.**

### 4.1.6 Grundsätze der Datenverarbeitung

Bei den allgemeinen Grundsätzen für die Verarbeitung personenbezogener Daten sind durch Art 5 DS-GVO keine signifikanten Änderungen im Vergleich zur geltenden Rechtslage ersichtlich (Verwendung nur nach Treu und Glauben und auf rechtmäßige Weise; Ermittlung nur für festgelegte, eindeutige und rechtmäßige Zwecke; Verwendung nur soweit für den Anwendungszweck wesentlich; Verwendung von sachlich richtigen Daten; Aufbewahrung nur solange erforderlich). Neu hinzugekommen ist durch die DS-GVO 2016/679 der Grundsatz der „Integrität und Vertraulichkeit“, welcher insb. für die rechtlichen Aspekte der technischen Verarbeitungsvorgänge und der Datensicherheit von Bedeutung ist. Der Verantwortliche muss

hiernach dafür sorgen, dass die Informationen vor unbefugter oder unrechtmäßiger Verarbeitung sowie vor Verlust, Zerstörung oder Schädigung geschützt werden.

Ein für die vorliegende Studie wesentliches Prinzip der Datenverarbeitung ist der **Zweckbindungsgrundsatz** (Art 5 Abs 1 lit b), der in weiterer Folge auch auf die rechtlichen Anforderungen der Einwilligungserklärung durchschlägt. Personenbezogene Daten dürfen nur für festgelegte, eindeutige und legitime Zwecke erhoben und nicht in einer mit diesen Zwecken unvereinbaren Weise weiterverarbeitet werden. Die Ausformulierung des Verarbeitungszwecks muss also erstens **eindeutig** sein, sodass kein Zweifel mehr bestehen bleiben darf. Art 12 Abs 1 iVm Art 13 Abs 1 lit c DS-GVO legen fest, dass die betroffenen Personen in präziser, transparenter und verständlicher Form sowie in einer klaren und einfachen Sprache über die Verarbeitungszwecke zu informieren sind („**Laienverständlichkeit**“, Plath, DSGVO$^2$ (2016) Art 5 Rz 6)

Dies schließt eine Berücksichtigung des Kontexts der konkreten Datenverarbeitung aus der Sicht eines kundigen Betrachters aber nicht aus. Unbestimmte „**Blankettformeln**“ sind dadurch aber ungültig, alle verwendeten Begriffe müssen stets im konkreten Kontext bestimmbar sein. Eine Einwilligung in die Datenverarbeitung zu „kommerziellen Zwecken“ wäre daher mangels Bestimmtheit ungültig. Die Festlegung ist hingegen ein formales Kriterium, welches erfordert, dass der materielle Gehalt des Zwecks auch in irgendeiner Form **festgelegt** wird. Schriftlichkeit ist hierbei kein generelles Erfordernis, auch wenn diese zu Beweiszwecken durchaus ratsam erscheint. Weiters ist der Zweck dann **legitim**, wenn die betroffene Person ihre Einwilligung zu einem festgelegten und eindeutigen Zweck gegeben hat oder eine andere gesetzliche Erlaubnis (zum Beispiel Erfüllung eines Vertrags gem. Art 6 Abs 1 lit b) vorliegt (Paal - Pauly, DS-GVO 2017 Art 5 Rz 27 f).

**Eine betroffene Person willigt immer nur in die Verarbeitung der sie betreffenden Informationen zu einem bestimmten Zweck ein. Dieser muss eindeutig und legitim sein (zum Beispiel die Datenverarbeitung für postalische Werbung), und in der Einwilligungserklärung für einen Laien verständlich ausformuliert werden. Eine nachträgliche Änderung des Zwecks lässt die zuvor abgegebene Einwilligungserklärung ungültig werden. Die genaue Analyse der rechtlichen Anforderungen an die Einwilligung erfolgt im nächsten Kapitel.**

Besonderes Augenmerk ist den **Grundsätzen der Datenminimierung und Speicherbegrenzung** (Art 5 Abs 1 lit c und e) zu schenken. Die DS-GVO 2016/679 sieht ausdrücklich vor, dass die Datenverarbeitung im Rahmen der Zweckbindung in qualitativer und quantitativer sowie zeitlicher Hinsicht zu begrenzen ist. Die DSGVO statuiert zunächst den anerkannten Grundsatz der Datenminimierung und gibt diesem – gegenüber dem geltenden DSG 2000 – eine klare sowie präzise Kontur. Gemäß Art 5 lit c DS-GVO müssen personenbezogene Daten „*dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein*“.

Eine Datenverarbeitung ist **dem Zweck angemessen**, wenn ihre Zuordnung zu den legitimierten Zwecken unstrittig ist. Zu prüfen ist demnach, ob die Datenverarbeitung erforderlich ist um diesen Zweck zu erreichen. Zudem muss die Datenverarbeitung dem Kriterium der **Erheblichkeit** entsprechen. Die Erheblichkeit schränkt die Verarbeitung der zunächst im Rahmen der Angemessenheit verarbeiteten Daten weiter ein. Eine dem Zweck angemessene Datenverarbeitung ist nicht notwendig auch erheblich; die Erheblichkeit ist mit Blick auf den Zweck nach objektiven Gesichtspunkten zu prüfen. Insofern ist nicht auf die subjektive Sichtweise des Verantwortlichen abzustellen. Wird der Erhalt eines Bewerbungsschreibens automatisch

erfasst, um die Einhaltung der Bewerbungsfrist zu überprüfen, ist dies dem Zweck angemessen. Sobald feststeht, dass die Frist gewahrt wurde und das Bewerbungsverfahren weitergeführt wird, ist diese Speicherung jedoch nicht mehr erheblich. Schließlich statuiert Art 5 lit c DS-GVO eine **Begrenzung der Datenverarbeitung auf das notwendige Maß**, wodurch gegenüber dem DSG 2000 und der DS-RL 95/46/EG, eine gewisse Verschärfung vorgenommen wurde. Wenngleich in aller Regel dieser Anforderung bereits durch das Erfüllen der Angemessenheit und Erheblichkeit hinreichend Rechnung getragen wird, bleibt ein nicht unbedeutender Anwendungsbereich bestehen. So können etwa Daten, deren bisherige Verarbeitung angemessen und erheblich war, im weiteren Verlauf ihre spezifische Relevanz mit Blick auf den legitimierten Zweck verlieren; ihre weitere Verarbeitung würde demnach über das „notwendige Maß" hinausgehen. Es geht dabei nicht um absolute Minimierung, sondern um Minimierung mit Augenmaß (Paal - Pauly, DS-GVO (2017) Art 5 Rz 35 ff; Plath, DSGVO$^2$ (2016), Art 5 Rz 10).

Eng mit dem Grundsatz der Datenminimierung verbunden ist der in Art 5 lit e DS-GVO statuierte **Grundsatz der Speicherbegrenzung**. Nach diesem Grundsatz besteht eine zeitliche Beschränkung, sodass die Daten nur so lange in einer Weise gespeichert werden dürfen, die eine Identifizierung der betroffenen Person gestattet, wie es der legitimierte Zweck erfordert. Damit besteht ein gewisser Rechtfertigungsdruck, wobei die Bestimmung gewisse (im interessierenden Zusammenhang wenig relevante) Ausnahmen enthält. Der Grundsatz hat deutliche Konsequenzen für die Weiterverarbeitung von Daten zu Sekundärzwecken, wenn diese zeitlich nach Erreichung des Primärzwecks erfolgt. Der Grundsatz greift jedoch nur insofern, als Daten in einer Form gespeichert werden, die eine Bestimmung der Identität der betroffenen Person – etwa durch Wiedervereinigung mit anderen Daten – gestattet (Paal - Pauly, DS-GVO 2017 Art 5 Rz 42).

**Die Verarbeitung ist umfänglich mit Blick auf den jeweiligen Zweck nicht bloß zu optimieren, sondern vielmehr zu minimieren. Dies schließt etwa redundantes Speichern personenbezogener Daten dann aus, wenn die Speicherung mit Blick auf den Verarbeitungszweck nicht angemessen und/oder nicht erforderlich ist oder die Relevanz in Hinblick auf den Verarbeitungszweck wegfällt. Dies erschwert eine Nutzung personenbezogener Daten im Rahmen von Big Data Analysen. Sind Datenverarbeitungen mit dem Grundsatz der Datenminimierung vereinbar, ergibt sich aus dem Grundsatz der Speicherbegrenzung eine zusätzliche Beschränkung in zeitlicher Hinsicht. Damit ist der Personenbezug gespeicherter Daten zu beseitigen, sobald dieser zur Erreichung des legitimierten Verarbeitungszwecks nicht mehr erforderlich ist. Den beschriebenen Grundsätzen ist im Rahmen der datenschutzfreundlichen Technikgestaltung (Privacy by Design, Art 25 DS-GVO) zwingend Rechnung zu tragen. So wird etwa dem Grundsatz der Datenminimierung durch den Einsatz von Single-Sign-on-Instrumenten in Hinblick auf Zugangsdaten und die Verwaltung von Stammdaten entsprochen. Bei der technischen Ausgestaltung sind im Rahmen einer Verhältnismäßigkeitsprüfung betriebswirtschaftliche Aspekte beachtlich, wobei diese eine extensive (den Grundsätzen der Datenminimierung und Speicherbegrenzung zuwiderlaufende) Datennutzung, etwa im Rahmen von Big Data, kaum rechtfertigen können.**

### 4.1.7 Zustimmung und Einwilligung

Zu Beginn eine terminologische Klarstellung: Die allgemeine **Datenschutzerklärung** eines Unternehmens gibt grundlegende Auskunft darüber, wie Daten unternehmensintern verwaltet und verarbeitet werden. Sie dient daher lediglich Informations- und allenfalls customer-relations Zwecken. Eine **Zustimmungserklärung** (Terminologie nach dem DSG 2000) oder **Einwilligung**

(nach der DS-GVO 2016/679) bezeichnet hingegen eine (standardisierte) rechtsgeschäftliche Erklärung, in der die Betroffenen im vornherein zustimmen, dass personenbezogene Informationen durch einen Verantwortlichen (Auftraggeber) verarbeitet werden dürfen.

#### 4.1.7.1 Zustimmung des Betroffenen

**DSG 2000:** Für eine Zustimmung zur Datenanwendung ist es gem. § 4 Z 14 DSG notwendig, dass der/die Betroffene diese für den jeweils konkreten Fall in Kenntnis der Sachlage und ohne Zwang abgibt. In einem solchen Fall sind die schutzwürdigen Geheimhaltungsinteressen des Betroffenen nicht berührt. Die Judikatur (siehe etwa OGH 20.03.2007, 4 Ob 221/06p; OGH 19.11.2002, 4 Ob 179/02f) entwickelte folgende Erfordernisse für Zustimmungsklauseln:

- Die Zustimmungsklausel ist im Text hervorzuheben.
- Bei einer Zustimmungserklärung müssen die Datenarten, der Übermittlungsempfänger sowie der Zweck beschrieben werden.
- Es muss auf den jederzeit möglichen Widerruf ausdrücklich hingewiesen werden.
- Die Schriftlichkeit der Zustimmung ist **nicht** notwendig.

Eine Datenverarbeitung ist gem. § 7 Abs 1 DSG nur zulässig, wenn Zweck und Inhalt der Datenanwendung von einer gesetzlichen Zuständigkeit oder einer rechtlichen Befugnis gedeckt sind und die schutzwürdigen Geheimhaltungsinteressen der betroffenen Personen gewahrt werden. Das DSG 2000 sieht die schutzwürdigen Geheimhaltungsinteressen des Betroffenen bei der Verwendung von nicht-sensiblen Daten als nicht verletzt, wenn ein Tatbestand des § 8 DSG erfüllt ist. Dies kann neben der Zustimmung des Betroffenen (Abs 1 Z 2) etwa bei berechtigten Interessen des Auftraggebers (Abs 1 Z 4) der Fall sein.

**DS-GVO 2016/679**: Durch die DS-GVO 2016/679 kommt es zu einer terminologischen Änderung von der „Zustimmungserklärung“ hin zur „Einwilligung“ der betroffenen Person. Auch nach Inkrafttreten der DS-GVO 2016/679 ist für das Datenschutzrecht das „**Verbotsprinzip**“ maßgebend, wonach die Verarbeitung personenbezogener Daten grundsätzlich verboten ist, sofern nicht ein Erlaubnistatbestand vorliegt (*Paal/Pauly*, DS-GVO 2017 Art 6 Rz 1).

In Art 6 Abs 1 DS-GVO werden die Fälle der rechtmäßigen Datenverarbeitung taxativ aufgezählt. Hiernach ist eine Verarbeitung nur rechtmäßig, wenn eine der Varianten erfüllt ist:

- Die betroffene Person hat ihre **Einwilligung zur Verarbeitung** gegeben (lit a; siehe dazu sogleich).
- Die Verarbeitung ist zur **Vertragserfüllung** oder zur Durchführung vorvertraglicher Maßnahmen notwendig, die auf Anfrage der betroffenen Person erfolgen (lit b).
- Die Verarbeitung ist zur Erfüllung rechtlicher Verpflichtungen des Verantwortlichen notwendig (lit c).
- Die Verarbeitung erfolgt zum Schutz lebenswichtiger Interessen einer natürlichen Person (lit d).
- Dem Verantwortlichen wurde eine Aufgabe im öffentlichen Interesse oder in Ausübung der öffentlichen Gewalt übertragen (lit e).

- Die Verarbeitung erfolgt zur Wahrung berechtigter Interessen des Verantwortlichen oder eines Dritten (lit f).

## 4.1.8 Rechtmäßige Datenverarbeitung

Das folgende Kapitel konzentriert sich insbesondere auf die Anforderungen, welche die DS-GVO 2016/679 an die rechtskonforme Erteilung einer Einwilligung zur Verarbeitung personenbezogener Daten stellt. Es ist uE empfehlenswert, dass die derzeit verwendeten Zustimmungserklärungen bereits an die Anforderungen der DS-GVO 2016/679 angeglichen werden, um die notwendigen Anpassungsmaßnahmen nach der nächstjährigen Gesetzesänderung so gering wie möglich zu halten.

### 4.1.8.1 Anforderungen an die Einwilligung

Die Einwilligung iSd Art 6 Abs 1 lit a DS-GVO wird in Art 4 Z 11 leg cit legaldefiniert. Hiernach muss die betroffene Person eine Willensbekundung freiwillig und für den bestimmten Fall, in informierter Weise und unmissverständlich abgeben. Dies kann entweder in Form einer Erklärung oder einer sonstigen eindeutig bestätigenden Handlung erfolgen, mit der die Person ihr Einverständnis zur Verarbeitung der personenbezogenen Daten zu verstehen gibt (ErwGr 171). Wenn sich eine Datenverarbeitung auf mehrere Zwecke bezieht, sollte für sie alle eine Einwilligung eingeholt werden (ErwGr 32). Die weiteren Anforderungen, welche die DS-GVO 2016/679 an eine wirksame Einwilligung stellt, finden sich in Art 7 DS-GVO.

Die DS-GVO 2016/679 verlangt bei schlichten personenbezogenen Daten keine ausdrückliche (oder gar schriftliche) Einwilligung, jedoch eine **unmissverständliche Abgabe der Willenserklärung**. Für die Gültigkeit ist es weiters unerheblich, ob die Willenserklärung auf Initiative der betroffenen Person abgegeben wurde. Ausreichend ist eine eindeutig bestätigende Handlung, die gem. Erwägungsgrund 32 etwa mündlich oder schriftlich erfolgen kann (Paal - Pauly, DS-GVO 2017 Art 6 Rz 11).

Beispielhaft aufgezählt werden folgende Möglichkeiten einer schriftlichen Einwilligung im Online-Bereich:

- Das Anklicken eines Kästchens auf einer Internetseite oder
- die Auswahl technischer Einstellungen für Dienste einer Informationsgesellschaft.

Auch andere Verhaltensweisen können als Willensbekundung geeignet sein, sofern die betroffene Person ihr Einverständnis zur Datenverarbeitung im jeweiligen Kontext signalisiert. Folgende Fälle **schließt** die DS-GVO 2016/679 in ErwGr 32 jedoch explizit als zulässige Einwilligung **aus**:

- Stillschweigen oder Untätigkeit der betroffenen Person oder
- bereits angekreuzte Kästchen auf einer Internetseite (Plath, DSGVO$^2$ (2016), § 4a BDSG Rz 50).

Die **Anforderungen** an eine gültige Einwilligung finden sich sowohl in Art 4 Z 11 DS-GVO als auch in Art 7 DS-GVO. Zusammengefasst muss eine Einwilligungserklärung anhand folgender Kriterien überprüft wurden (*Plath*, DSGVO$^2$ (2016), Art 7 Rz 2):

- Liegt eine unmissverständliche Erklärung oder sonstige eindeutig bestätigende Handlung der betroffenen Person vor? (Art 4 Z 11 DS-GVO)

- Hat die betroffene Person Kenntnis über die Sachlage? (Art 4 Z 11 DS-GVO)
- Wurde die Einwilligung ohne Zwang abgegeben, insb. ohne rechtswidrige Koppelung? (Art 4 Z 11 DS-GVO, Art 7 Abs 4 DS-GVO)
- Wurde die Einwilligung für einen bestimmten Fall abgegeben? (Art 4 Z 11 DS-GVO)
- Ist die Abgabe der Einwilligung nachweisbar? (Art 7 Abs 1 DS-GVO)
- Falls es sich um eine „kombinierte" Erklärung (zum Beispiel in Allgemeinen Geschäftsbedingungen) handelt: Kann die Einwilligung von dem übrigen Text leicht unterschieden werden? (Art 7 Abs 2 DS-GVO)
- Liegt ein Widerruf der Einwilligung vor? (Art 7 Abs 3 DS-GVO)

Die DS-GVO 2016/679 sieht vor, dass ein Ersuchen um Einwilligung in verständlicher und leicht zugänglicher Form und in einer klaren und einfachen Sprache erfolgen muss. Empfehlenswert – wenn auch nicht gesetzlich verpflichtend – ist die Verwendung des Wortes „Einwilligung" sowie eine grafische Hervorhebung. Die Klauseln müssen neben den Anforderungen des Art 4 Z 11 DS-GVO den **Verantwortlichen** nennen sowie den **Zweck** der Verarbeitung von personenbezogenen Informationen, um eine Einwilligung „in informierter Weise" zu gewährleisten (ErwGr 42; Albrecht - Jotzo, (2016), 71).

Der Text muss daher so ausgestaltet sein, dass die Einwilligung leicht wahrgenommen werden kann, was der Warn- und Hinweisfunktion geschuldet ist. Es soll dadurch vermieden werden, dass bei komplexen schriftlichen Erklärungen eine Einwilligung nur beiläufig erteilt, oder aber die Einwilligung im „Kleingedruckten" versteckt wird. Ein Verstoß gegen die Vorschriften der DS-GVO führt gem. Art 7 Abs 2 Satz 2 dazu, dass die jeweiligen Teile der Erklärung nicht rechtsverbindlich werden, während jedoch die anderen Vertragsbedingungen davon nicht betroffen sind (Plath, DSGVO$^2$ (2016) Art 7 Rz 9).

Sehr wohl kann eine datenschutzrechtliche Zustimmungserklärung auch als Teil von vorformulierten Vertragsbedingungen wirksam eingeholt werden. Bedeutung erlangt dies vor allem bei der Verwendung von standardisierten **Allgemeinen Geschäftsbedingungen (AGB)**. Wenn die Zustimmung zur Datenverarbeitung gleichzeitig mit den AGB Vertragsbestandteil wird, unterliegt sie der AGB-Kontrolle nach dem Allgemeinen Bürgerlichen Gesetzbuch und dem Konsumentenschutzgesetz (Jahnel, (2010), 3/142).

Eine Einwilligung ist gem. Art 4 Z 11 DS-GVO nur dann gültig, wenn sie freiwillig abgegeben wird. **Freiwilligkeit** bedeutet, dass tatsächlich eine Wahl für die betroffene Person besteht. Hierfür ist es erforderlich, dass die Einwilligung verweigert oder auch zurückgezogen werden kann, ohne dass dadurch Nachteile entstehen (ErwGr 42). Dies wird durch das „**Kopplungsverbot**" gem. Art 7 Abs 4 DS-GVO konkretisiert: Die Einwilligung in eine Datenverarbeitung ist dann nicht mehr freiwillig, wenn die betroffene Person ein Produkt nicht erwerben oder eine Dienstleistung nicht nutzen kann, ohne in die Verarbeitung von Informationen einzuwilligen, die zur Erfüllung des Vertrags gar **nicht erforderlich** sind. Es ist strittig, ob das Kopplungsverbot nur im Fall einer Monopolstellung gilt, das heißt wenn es nur einen Anbieter der Leistung gibt und die betroffene Person diese ohne Preisgabe ihrer Daten nicht erhält (Plath, DSGVO$^2$ (2016) Art 7 Rz 14).

Die betroffene Person hat jederzeit das Recht, ihre zuvor gegebene Einwilligung ohne Angabe von Gründen zu widerrufen (Art 7 Abs 3 DS-GVO). Der Widerruf muss ebenso **einfach** erfolgen können wie die vorherige Abgabe der Willenserklärung. Das bedeutet zum Beispiel, dass eine

mündliche erteilte Einwilligung auch wieder mündlich gegenüber dem/der Verantwortlichen widerrufen werden kann, und nicht etwa die Schriftform verlangt werden darf. Die betroffene Person ist von den Rechten und der Wirkung einer Widerrufserklärung bereits vor Abgabe der Einwilligung zu informieren.

**Hervorzuheben ist, dass bereits bestehende Datenverarbeitungen bis Mai 2018 mit der Verordnung in Einklang zu bringen sind. Wenn eine Zustimmung nach dem DSG 2000 für eine solche Datenverarbeitung vorliegt, müssen die Kunden nicht erneut ihre Einwilligung erteilen, sofern die zuvor erteilte Zustimmungserklärung auch den Anforderungen der DS-GVO 2016/679 entspricht (ErwGr 171). Es ist für Unternehmen jedoch nicht ohne weiteres ersichtlich, ob ihre derzeit verwendete Zustimmungserklärung mit der DS-GVO 2016/679 in Einklang stehen wird oder ob erneut eine Einwilligung aller betroffenen Personen eingeholt werden muss. Dies kann negative Auswirkungen auf Innovationsprozesse haben.**

### 4.1.9 Rechtliche Aspekte technischer Verarbeitungsvorgänge und der Datensicherheit

Zur Gewährleistung eines hohen Schutzniveaus für personenbezogene Informationen gehört auch Datenschutz durch Technikgestaltung sowie durch datenschutzfreundliche Voreinstellungen. Einleitend ist festzuhalten, dass sich hinsichtlich des sicherzustellenden technisch-organisatorischen Schutzniveaus das DSG 2000 und die DS-GVO 2016/679 nur graduell unterscheiden. Auch weiterhin ist ein **Höchstmaß an Datensicherheit** zu gewährleisten, soweit Kosten und Nutzen in einem angemessenen Verhältnis stehen.

**DSG 2000:** Alle Organisationseinheiten eines Auftraggebers oder Dienstleisters sind gemäß § 14 DSG verpflichtet, Maßnahmen zur Gewährleistung der **Datensicherheit** (iSv Vertraulichkeit, Integrität und Verfügbarkeit) zu treffen. Sicherzustellen ist, dass die Daten vor zufälliger oder unrechtmäßiger Zerstörung und vor Verlust geschützt sind, dass ihre Verwendung ordnungsgemäß erfolgt und dass die Daten Unbefugten nicht zugänglich sind. Im Einzelnen sind die dafür geeigneten Maßnahmen nach Maßgabe der Art der verwendeten Daten, nach Umfang und Zweck der Verwendung sowie unter Bedachtnahme des Stands der technischen Möglichkeiten und der wirtschaftlichen Vertretbarkeit zu treffen. Mit anderen Worten: Technisch-organisatorische Datensicherheit nach dem Stand der Technik, aber nicht um jeden Preis.

Das Gesetz enthält in § 14 Abs 2 DSG eine Aufzählung von Mindestanforderungen, wie die Festlegung einer datenschutzrechtlichen Aufbau- und Ablauforganisation, betriebsinterne Belehrungspflichten gegenüber MitarbeiterInnen, Zutritts- und Zugriffskontrollen sowie Protokollführung und Dokumentation. Die „**wirtschaftliche Vertretbarkeit**“ wird dahin konkretisiert, dass unter Berücksichtigung des Stands der Technik und der daraus resultierenden Kosten ein „angemessenes“ Schutzniveau sicherzustellen ist. Aus dem DSG 2000 ergeben sich keine weiteren Spezifizierungen hinsichtlich der Datensicherheit (Dohr – Pollirer – Weiss - Knyrim, DSG$^2$ (2015) § 14).

**DS-GVO 2016/679:** Bereits Art 5 DS-GVO setzt die Integrität und Vertraulichkeit der Datenverarbeitung voraus und statuiert, dass eine angemessene Sicherheit der personenbezogenen Daten durch geeignete **technische und organisatorische Maßnahmen** zu gewährleisten ist. Dies beinhaltet insbesondere den Schutz vor unbefugter oder unrechtmäßiger Verarbeitung und vor unbeabsichtigtem Verlust, unbeabsichtigter Zerstörung oder unbeabsichtigter Schädigung. Konkretisierend verlangt Art 25 DS-GVO eine

datenschutzfreundliche Technikgestaltung (**Privacy by Design;** siehe weiterführend Wolff - Brink, BeckOK Datenschutzrecht[18] § 3a BDSG Rz 60 ff).

Demgemäß sind geeignete technische und organisatorische Maßnahmen zu treffen, die dafür ausgelegt sind, die **Datenschutzgrundsätze** (Art 5 DS-GVO) wirksam umzusetzen und die notwendigen Garantien in die Verarbeitung aufzunehmen, um den Anforderungen der DS-GVO 2016/679 zu genügen und die Rechte der betroffenen Personen zu schützen. Dies muss unter Berücksichtigung des Stands der Technik, der Implementierungskosten und der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung sowie der unterschiedlichen Eintrittswahrscheinlichkeit und Schwere der mit der Verarbeitung verbundenen Risiken für die Rechte und Freiheiten natürlicher Personen vom Verantwortlichen sowohl zum Zeitpunkt der Festlegung der Mittel für die Verarbeitung als auch zum Zeitpunkt der eigentlichen Verarbeitung erfolgen (Art 25 Abs 1 DS-GVO). Außerdem dürfen nur solche personenbezogenen Daten verarbeitet werden, die für die Erreichung des jeweiligen Zwecks erforderlich sind. Diesbezüglich sind geeignete technische und organisatorische Voreinstellungen zu treffen (Art 25 Abs 2 DS-GVO).

Daran schließt die – auch den Auftragsverarbeiter gemäß Art 28 Abs 3 lit c DS-GVO treffende – Zentralnorm betreffend Datensicherheit an: Art 32 DS-GVO statuiert, dass der **Verantwortliche oder Auftragsverarbeiter** geeignete technische und organisatorische Maßnahmen treffen muss, um ein dem Risiko **angemessenes Schutzniveau** zu gewährleisten. Diese Maßnahmen schließen insbesondere Folgendes ein:

- Pseudonymisierung und Verschlüsselung personenbezogener Daten (Art 32 Abs 1 lit a DS-GVO)
- Dauerhafte Vertraulichkeit, Integrität, Verfügbarkeit und Belastbarkeit der Systeme und Dienste (Art 32 Abs 1 lit b DS-GVO)
- Rasche Wiederherstellung der Verfügbarkeit und des Zugangs zu den Daten nach einem technischen Zwischenfall (Art 32 Abs 1 lit c DS-GVO)
- Überprüfung, Bewertung und Evaluierung der technischen und organisatorischen Maßnahmen (Art 32 Abs 1 lit d DS-GVO)

Demgemäß müssen Verantwortliche oder Auftragsverarbeiter die Risiken der jeweiligen Datenverarbeitung reflektieren und verhältnismäßige risikoadäquate Maßnahmen ergreifen. Im Rahmen der Verhältnismäßigkeit sind die Implementierungskosten den drohenden Risiken und technischen Möglichkeiten gegenüber zu stellen. Implementierungskosten sind Kosten, die ein Verarbeiter aufwenden muss, um eine Datensicherheitsmaßnahme in seinem Verarbeitungssystem zu integrieren. Folgekosten (im laufenden Betrieb) sind – anders als nach Art 17 Abs 1 DSRL – nicht erfasst (Paal - Pauly, DS-GVO 2017 Art 32 Rz 3, 60).

Hervorzuheben ist die Zielvorgabe der Sicherung von **Integrität, Vertraulichkeit, Belastbarkeit und Verfügbarkeit der Systeme und Dienste**. Der Verordnungsgeber hat damit fundamentale Aspekte moderner IT-Sicherheit adressiert, lässt aber freilich im Einzelnen offen, durch welche Maßnahmen diese Ziele erreicht werden können. Die Integrität lässt sich beispielsweise durch elektronische Signaturen oder Eingabekontrollen sicherstellen. Das Kriterium der Verfügbarkeit stellt darauf ab, dass eine zufällige Zerstörung oder ein zufälliger Verlust durch geeignete Maßnahmen, insbesondere Back-ups, verhindert wird. Im Fall von Sicherungen wird auf den Grundsatz der Datensparsamkeit entsprechend Bedacht zu nehmen sein. Das Kriterium der Belastbarkeit bezieht sich auf die Widerstandskraft oder Leistungsfähigkeit des

Datenverarbeitungssystems im Umgang mit Attacken Dritter, insbesondere „Denial-of-Service"-Attacken.

Die DS-GVO 2016/679 legt kein absolutes statisches **Sicherheitsniveau** fest. Vielmehr zielt die Verordnung technologieneutral darauf ab, dass ein dem jeweiligen technologischen Umfeld dynamisch-angepasstes angemessenes Schutzniveau aufrechterhalten wird. Ausgangspunkt ist dabei eine Bewertung möglicher (vorhersehbarer) Risiken, denen der Auftragsverarbeiter potentiell ausgesetzt ist, des technisch Möglichen sowie des wirtschaftlich Machbaren. Diese Bewertung fließt schließlich in eine Gesamtabwägung ein, die in einem konkreten Datensicherheitskonzept zu münden hat. Hinsichtlich der Risiken ist freilich zu betonen, dass, je sensibler Daten sind, desto höher das mit der Verarbeitung verbundene Risikoniveau ist (Paal - Pauly, DS-GVO 2017, Art 32 Rz 36, 39, 53).

**Zusammengefasst kann festgehalten werden, dass die Verarbeitung personenbezogener Daten stets höchsten technischen Standards zu entsprechen hat. Das konkrete Sicherheitsniveau und die jeweiligen technisch-organisatorischen Maßnahmen sind auf Grundlage einer Risiko- und Bedrohungsanalyse sowie einer Evaluierung des technologischen Umfelds vorzunehmen, wobei die Art der verarbeiteten Daten (zum Beispiel Gesundheitsdaten) ein hohes Risiko und demgemäß ein hohes Sicherheitsniveau nahelegen können. Dies gilt insbesondere hinsichtlich Angriffe Dritter, unberechtigter Zugriffe durch Mitarbeiter oder fehlerhafter Verarbeitungsvorgänge. Datenverarbeitende Unternehmen müssen die (zeit- und kostenintensive) Analyse ihrer konkreten Risiko- und Bedrohungssituation vor Inkrafttreten der DS-GVO abschließen, um die erforderlichen technischen Lösungen vor Mai 2018 implementieren zu können.**

### 4.1.9.1 Bewertung rechtlicher Sonderaspekte von Big Data Anwendungen

Im Folgenden werden einige Sonderaspekte dargestellt, die für die rechtliche Beurteilung des gegenständlichen Use Case von besonderer Relevanz sind und signifikante Auswirkungen auf innovationstreibende Unternehmen haben können.

### 4.1.9.2 Einwilligung zur Datenverarbeitung für Testzwecke

Für die Entwicklung neuer Dienstleistungen oder Produktfeatures kann es notwendig sein, personenbezogene Daten, die zuvor rechtmäßig ermittelt wurden, für die Testphase dieser neuen Features zu nutzen. Hierbei handelt es sich uE nicht um eine legitime Zweckänderung iSv Art 6 Abs 4 DS-GVO, sondern wiederum um eine Datenverarbeitung, die den datenschutzrechtlichen Anforderungen der DS-GVO 2016/679 entsprechen muss. Der datenschutzrechtlich Verantwortliche muss daher eine Einwilligung für eine Datenverarbeitung zu Testzwecken von den betroffenen Personen einholen. Die Testzwecke müssen möglichst konkret ausformuliert werden, um eine informierte Einwilligung der betroffenen Person für einen bestimmten Fall zu ermöglichen (Art 4 Z 11 DS-GVO). Der alleinige Verweis auf eine Datenverarbeitung „zu Testzwecken" ist uE hierfür nicht ausreichend, da zu wenig bestimmt.

**Dies stellt eine bedeutende Erschwerung im Produktentwicklungsprozess dar, da man eine erneute Einwilligung benötigt, bevor die – bereits im Unternehmen vorhandenen – personenbezogenen Informationen für den Test eines neuen Produkts verwendet werden dürfen. Als Ausweg bietet sich an, dass die Mitglieder des Produktentwicklungsteams ihre eigenen Daten verwenden. Auch wird man bei Verwendung moderner Innovationskonzepte für die Erhebung und den Test der Kundenanforderungen einen**

**Kreis von Testanwendern als Innovationspartner akquirieren, von denen dann auch eine entsprechende datenschutzrechtliche Einwilligung erlangt werden kann.**

#### 4.1.9.3 Opt-out

Eine Einwilligung kann nur mittels eindeutig bestätigender Handlung abgegeben werden. Die betroffene Person muss einen separaten Textabschnitt unterschreiben oder – im Online-Bereich – ein deutlich abgesetztes Kästchen ankreuzen. Nicht zulässig ist es, dass bei einer elektronisch abzugebenden Einwilligung ein entsprechendes Häkchen bereits gesetzt wird und die betroffene Person dieses entfernen muss, wenn sie der Datenverarbeitung nicht zustimmen möchte. Gemäß ErwGr 32 kann eine gültige Einwilligung weder durch Stillschweigen oder Untätigkeit der betroffenen Person erfolgen, noch durch bereits angekreuzte Kästchen auf einer Internetseite. Diese Diskussion wird unter dem Schlagwort **„opt-out“** geführt (Paal - Pauly, DS-GVO (2017), Art 7 Rz 13, 15).

In Deutschland soll es nach derzeit hL und RSP zum BDSG jedoch ausreichen, wenn die Einverständniserklärung mittels opt-out unterlassen werden kann. Das bedeutet, dass ein bereits gesetztes Häkchen auf einer Internetseite entfernt werden muss, sofern man mit der Datenverarbeitung nicht einverstanden ist. Dies kann jedoch längstens bis zum Inkrafttreten der DS-GVO 2016/679 im Mai 2018 gelten, da ErwGr 32 ebendieses Vorgehen für unzulässig erklärt. Ein Verstoß gegen die Vorschriften der DS-GVO führt gem. Art 7 Abs 2 Satz 2 dazu, dass die jeweiligen Teile der Erklärung nicht rechtsverbindlich werden, während jedoch die anderen Vertragsbedingungen davon nicht betroffen sind (Plath, DSGVO$^2$ (2016), Art 7 DSGVO Rz 9).

Zur Klarstellung muss folgende Sachverhaltskonstellation abgegrenzt werden: Es mag aus betriebswirtschaftlichen Gründen vorteilhaft sein, Kunden neue Produkte oder Produkterweiterungen anhand der sie betreffenden Echtdaten vorzuführen (z.B. Versicherungsangebote, etc.). Hierfür verarbeitet ein Unternehmen personenbezogene Informationen und erstellt daraus ein individualisiertes Produkt oder eine individualisierte Dienstleistung. Diese sollen dann zu Marketingzwecken eingesetzt werden und zu besseren Werbeergebnissen führen.

Werbung mit individualisierten Leistungsangeboten erfolgt häufig im Online-Bereich, wobei dem Kunden regelmäßig die Möglichkeit eingeräumt wird, das Produkt- oder Dienstleistungsangebots per Mouse-Click abzulehnen. Wenn jedoch bereits für die Erstellung des individualisierten Leistungsangebots personenbezogene Daten verarbeitet werden, unterliegt diese Datenverarbeitung vollumfänglich den datenschutzrechtlichen Bestimmungen und darf daher nicht ohne Einwilligung des Betroffenen erfolgen. Als Rechtfertigung reicht es nicht aus, dass der potentielle Kunde das Leistungsangebot ablehnen und auch die Datenverarbeitung für die Zukunft untersagen kann, da eine einmalige Verarbeitung seiner personenbezogenen Informationen bereits erfolgte.

**Bei der Illustration neuer Features anhand der Echtdaten von Kunden werden personenbezogene Informationen verarbeitet. Diese Verarbeitung darf nicht – auch nicht zur einmaligen Veranschaulichung – zustimmungslos erfolgen. Die Möglichkeit, die bereits erfolgte Datenverarbeitung für die Zukunft zu untersagen, legitimiert die vorangehende Datenverarbeitung nicht.**

### 4.1.9.4 Einwilligung zur Datenverarbeitung im Rahmen von Big Data

Um das Phänomen „Big Data“ rechtlich beurteilen zu können, bedarf es zunächst eines einheitlichen Begriffsverständnisses. Big Data ist jedoch weder in der DS-GVO definiert, noch liegt ein allgemeingültiges Begriffsverständnis vor. Der Grundgedanke von Big Data ist, möglichst viele Informationen aus möglichst vielen verschiedenen Quellen zusammenzuführen und anschließend mittels algorithmischer Verfahren zu analysieren. Der Begriff „Big Data“ wurde maßgeblich dadurch geprägt, dass traditionelle Methoden der Datenverarbeitung aufgrund der Größe der Datenmenge scheitern. Neue Technologien (zum Beispiel Map Reduce oder die entsprechende Open-Source Version Hadoop) ermöglichen nun eine Analyse, indem wenig oder schwach strukturierte Daten kombiniert werden. Die Besonderheit besteht darin, dass Daten, die a priori keinen Bezug zueinander aufweisen, auf dezentrale Speicherorte gelegt und parallel bearbeitet werden. Das Analyseergebnis erhält man erst aus der Verknüpfung dieser Daten (Dohr – Pollirer – Weiss - Knyrim, DSG$^2$ (2015) § 6).

Ein Wesensmerkmal von Big Data ist, dass große Informationsmengen zusammengeführt werden, ohne den vormaligen Ermittlungs- oder Verwendungszweck zu berücksichtigen. Der neue Analysezweck für die Weiterverwendung der Daten wird erst nachträglich (also nach der Ermittlung) festgelegt. Wenn Daten zu einem anderen als dem ursprünglichen Ermittlungszweck verwendet werden, liegt aus datenschutzrechtlicher Sicht eine Datenübermittlung vor, die wiederum den Anforderungen der DS-GVO zu genügen hat. Big Data soll in der Praxis aber gerade ermöglichen, den jeweiligen Analysezweck erst nach der Datenaggregation festzulegen. Dies stellt einen Widerspruch zu dem allgemeinen Grundsatz der **Zweckbindung** gem. § 6 Abs 1 Z 2 DSG dar (Jahnel, (2016)).

Die Frage nach einer gültigen Zustimmung des Betroffenen (§ 4 Z 14 DSG) eröffnet zwei Probleme: Zum einen ist eine generelle Zustimmung einer Person zu zukünftigen (noch nicht bekannten) Datenverwendungen ungültig, da eine solche nicht in Kenntnis der Sachlage und für den konkreten Fall abgegeben wird. Eine Zustimmungserklärung, die von einer betroffenen Person **im Vorfeld für Big Data** Anwendungen abgegeben wird, kann daher **nie gültig** sein. Eine solche müsste vielmehr erst dann erfolgen, wenn der konkret gewünschte Analysezweck feststeht. Dies ist aber im Bereich von Big Data unpraktikabel, da der Betroffene hierfür erst aus dem riesigen Datenkonvolut identifiziert werden muss, um anschließend seine Einwilligung zur Weiterverwendung einzuholen.

Das Speichern von Informationen auf Vorrat, um diese für eine allfällige spätere Analyse aufzubewahren, kann weiters nicht mit dem Grundsatz der **Datensparsamkeit** gem. § 6 Abs 1 Z 3 DSG vereinbart werden. Hiernach dürfen nur solche Informationen verwendet werden, die für den Zweck der jeweiligen Datenanwendung wesentlich sind und darüber nicht hinausgehen.

**Die rechtlichen Anforderungen an die Einwilligung der betroffenen Person sind angesichts der empirisch gut abgesichert Bedeutung von „Defaults“ von erheblicher Bedeutung für die Adoption von Innovationen. Dieses Phänomen besagt, dass durch die Voreinstellung einer bestimmten Auswahl deren Akzeptanz erhöht wird. Kombiniert mit dem oben beschriebenen Verbot der Demonstration des Wertes des neuen Dienstes auf Basis bestehender Daten auf Basis von Verarbeitungen, zu denen keine Einwilligung besteht, ist dies ein wesentliches Hindernis für die Big Data getriebene Innovation. Prima facie könnte man daher zur Schlussfolgerung kommen, dass durch die DS-GVO die Anwendung von Big Data bei personenbezogenen Daten verhindert wird. Allerdings lassen sich durch Kombination mit technischen Verfahren und geeigneten Vorgangsweisen zur Erlangung**

**von Einwilligungen Big Data Anwendungen möglich, die auch den legitimen Schutzbedürfnissen der Anwender entsprechen. Diese werden in den Empfehlungen in Kapitel 6 beschrieben.**

#### 4.1.9.5 Profiling

Der Begriff des „Profiling“ wird in der DS-GVO in Art 4 Z 4 definiert: Hierunter versteht man eine automatisierte Verarbeitung von personenbezogenen Daten, um bestimmte persönliche Aspekte einer Person zu bewerten. Insbesondere geht es dabei um die Analyse oder **Vorhersage** persönlicher Informationen, wie etwa Arbeitsleistung, wirtschaftliche Lage, Gesundheit, persönliche Vorlieben, Verhalten, etc.

Für das Profiling wurden verschiedene Analysemethoden entwickelt, zum Beispiel das Tracking (Eigenschaften und Verhalten iZm dem Aufenthaltsort werden aufgezeichnet) oder das Scoring (individuelle Eigenschaften werden bewertet). Profiling wird unter anderem dazu verwendet, Persönlichkeitsprofile zu erstellen und damit Angebote zu personalisieren. Eine weitere Anwendungsmöglichkeit besteht im sogenannten Geomarketing: Mobilfunkdaten sowie Daten aus dem öffentlichen WiFi können analysiert werden, um Verkehrsströme zu ermitteln und mit soziodemografischen Daten zu verknüpfen. Diese Informationen können dann für Entscheidungen in der Stadtentwicklung herangezogen werden.

Als wichtige Einschränkung beim Profiling gilt § 22 Abs 1 DS-GVO, der besagt, dass eine Person grundsätzlich nicht allein aufgrund einer automatisierten Datenverarbeitung einer Entscheidung unterworfen werden darf, die eine **rechtliche Wirkung ihr gegenüber** entfaltet oder sie sonst beeinträchtigt (zum Beispiel die automatische Ablehnung eines Online-Kreditantrages; ErwGr 71).

**Durch eine Analyse der Big Data Datenbestände soll das Verhalten einer bestimmten Person vorausgesagt werden. Für diese Art der Datenverarbeitung sind die datenschutzrechtlichen Sonderbestimmungen für Profiling zu beachten (Art 22 DS-GVO), sofern damit rechtliche Wirkungen verbunden sind. Die Anzeige einer personalisierten Werbung etwa entfaltet diese nicht.**

## 4.2 ePrivacy

### 4.2.1 Überblick über sektorspezifische Regulierungen (TKG)

#### 4.2.1.1 Historische Entwicklung

Die ersten Arbeiten zum Thema Datenschutz und Telekommunikation fanden im Rahmen des Europarates Mitte der 1980er Jahre statt und mündeten in der Empfehlung R (95) 4 des Ministerkomitees des Europarates vom 7. Februar 1995 über den Schutz personenbezogener Daten im Bereich der Telekommunikationsdienste unter besonderer Berücksichtigung von Telefondiensten. Die Vorarbeiten dazu flossen bereits 1993 in das damalige österreichische Fernmeldegesetz 1993 (BGBl 1993/908) ein und bildeten einen eigenen Datenschutzteil in diesem Gesetz.

#### 4.2.1.2 Geltende Regelungen

Die zentralen Fragen des Datenschutzes im Fernmeldegesetz 1993 wurden auch ins Telekommunikationsgesetz 2003 (BGBl. I 2003/70) übernommen und bilden dessen 12. Abschnitt. Abgesehen von größeren Veränderungen im Zusammenhang mit der Einführung und der darauffolgenden Aufhebung der Vorratsdatenspeicherung durch den Verfassungsgerichtshof, sind die für die vorliegende Fragestellung relevanten Regelungen im Wesentlichen den ursprünglichen Vorschriften ähnlich. Es kann nicht geleugnet werden, dass die Regelungen sehr telefonlastig sind und die Fragestellungen moderner Technologien, aus denen sich neue Dienste, insbesondere OTT entwickelt haben, unbeantwortet bleiben müssen.

#### 4.2.1.3 Zentrale Fragestellungen

Mit der Entwicklung insbesondere der Internettelefonie haben sich unzählige Fragen gestellt. Ein zentraler Punkt ist das Problem des Wettbewerbs, nämlich inwieweit für Dienste, die im Erlebnis für den Nutzer identisch mit einem klassischen Telefondienst sind, gleichartige wettbewerbsrechtliche Rahmenbedingungen herrschen sollen oder müssen. Der klassische Ansatz würde solchen Diensten, wie etwa Skype, die Berechtigung absprechen können, weil manche zentralen Elemente eines Telefondienstes nicht geboten werden können, etwa der Zugang zu Notrufdiensten, welche auch zur jeweils in Frage kommenden Einsatzzentrale geroutet werden. Gerade die Frage der Standortbestimmung in solchen Fällen ist nicht ohne Relevanz.

Wesentlich aktueller sind jedoch die Fragen im Zusammenhang mit der Handhabung jener Daten, die im Zuge der Erbringung des Dienstes anfallen. Besonders bewusst wurde dies der breiten Öffentlichkeit mit der Einführung der (mittlerweile wieder aufgehobenen) Vorratsdatenspeicherung, als in § 102a TKG 2003 eine sehr umfangreiche Liste von jenen Daten, die im Zuge der Diensterbringung anfallen, als Speicherpflichtung definiert wurde und diese Liste sich zusätzlich auf Telefon- und Internetdaten bezogen hat.

#### 4.2.1.4 Datenverwendung nach geltendem Recht

Das eigentliche Problem sind jedoch weniger jene Daten, die bei der Erbringung eines klassischen Telefon- oder Internetdienstes anfallen, sondern jene Daten, die bewusst gesammelt werden, um Werbung oder andere kommerzielle Ziele zu verfolgen, womit aber auch der Dienst kostenlos angeboten werden kann. So hat es einen Grund, warum Facebook von seinen Usern die Telefonnummer abverlangt, auch wenn dies nicht Voraussetzung des Dienstes ist. Durch Verknüpfung von Daten aus Facebook- und WhatsApp etwa, entstehen wertvolle Persönlichkeitsprofile, welche sich gut vermarkten lassen, die jedoch auch eine extreme Gefahr darstellen können.

Die zentrale Regelung für alle Fragen, die die aus dem Dienst gewonnenen Daten betrifft, ist die Handhabung von Verkehrsdaten. Unter Verkehrsdaten versteht das TKG jene Daten, die zum Zwecke der Weiterleitung einer Nachricht an ein Kommunikationsnetz oder zum Zwecke der Fakturierung dieses Vorgangs verarbeitet werden. Darunter fallen auch Zugangsdaten, die es ermöglichen, dynamische IP-Adressen an Hand eines vorgegebenen Zeitpunktes einem Nutzer zuzuordnen.

§ 99 TKG 2003 legt für Verkehrsdaten zwei Grundprinzipien fest.

#### 4.2.1.4.1 Datenlöschung nach der Kommunikation

§ 99 Abs. 1 TKG 2003 schreibt vor, dass Verkehrsdaten grundsätzlich nicht gespeichert oder übermittelt werden dürfen und vom Betreiber nach Beendigung der Verbindung zu löschen sind. Ausnahmen davon gelten für Verrechnungszwecke oder – wie dies im Fall der Vorratsdatenspeicherung der Fall war - aufgrund einer spezifischen gesetzlichen Anordnung.

#### 4.2.1.4.2 Verbot der Auswertung der Daten

§ 99 Abs 4 TKG beschränkt die Auswertung der Daten eines Teilnehmeranschlusses auf die Zwecke der Verrechnung, es sei denn, der Teilnehmer stimmt einer darüberhinausgehenden Verwendung zu, die aber wiederum ausschließlich zum Zwecke der Vermarktung der eigenen Telekommunikationsdienste zulässig wäre. Eine Zustimmung zur Verwendung der individualisierten personenbezogenen Daten wäre angesichts dieser Bestimmung zivilrechtlich unwirksam.

#### 4.2.1.4.3 Datenhandhabung bei Anonymisierung

Diese beiden Prinzipien stammen erkennbar aus der Telefonwelt, wo einzeln verrechenbare Kommunikationsvorgänge klar voneinander unterscheidbar sind und daher die Daten strukturiert behandelt werden können. Für die rechtliche Einordnung kommerzieller Auswertungen, wie dies von großen Internetdiensten vorgenommen werden, sind diese Regelungen ungeeignet.

Um diese Anforderungen zu erfüllen, ist es unerlässlich, dass die Daten so anonymisiert werden, dass sie nicht oder mit hoher Wahrscheinlichkeit nicht Rückschlüsse auf vom Datenschutzgesetz geschützte Personenkreise zulässt. Damit verlieren die Daten ihre Eigenschaft als personenbezogene Daten und verlassen den Anwendungsbereich des TKG 2003. Das TKG behandelt allerdings nicht die sogenannte „Pseudonymisierung“, also die nach der Definition der Datenschutz-Grundverordnung erfolgende Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden. Diese Daten werden, weil sie mit Wahrscheinlichkeit identifizierbar sind, wohl bis zum Inkrafttreten der Datenschutz-Grundverordnung vermutlich jedoch weiterhin als personenbezogene Daten anzusehen sein.

Man darf zwar nicht außer Acht lassen, dass der Schutz von juristischen Personen in Österreich spätestens mit Inkrafttreten der Datenschutz-Grundverordnung endet, allerdings wird sich die Frage der Auswertung von Unternehmensanschlüssen auch weiterhin stellen, weil oder soweit diese individuell bestimmbaren Mitarbeitern zuzuordnen sind, etwa bei Firmenhandys.

Es wird daher – auch nach der künftigen Rechtslage – zumindest so lange, bis zur neuen Rechtslage ausreichend Judikatur vorliegt, kein Weg daran vorbeiführen, für die Nutzung von Big Data einen Anonymisierungsgrad zu finden, der jede Rückführbarkeit und jede Identifizierbarkeit tatsächlich ausschließt.

## 4.2.2 Aktuelle Entwicklung auf europäischer Ebene

### 4.2.2.1 Geltende Regelungen

Derzeit sind die telekommunikationsspezifischen datenschutzrechtlichen Regelungen in der Richtlinie 2002/58/EG des europäischen Parlaments und des Rates vom 12. Juli 2002 über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation - Datenschutzrichtlinie für elektronische Kommunikation enthalten. Die letzte Änderung erfolgte durch die Richtlinie 2009/136/EG.

Die Umsetzung in nationales Recht erfolgte in Österreich im 12. Abschnitt des TKG 2003.

### 4.2.2.2 E-Privacy

Die Europäische Kommission hat am 16. Jänner 2017 als sektorenspezifische Ergänzung der Datenschutz-Grundverordnung den Entwurf für die e-Privacy Verordnung (Vorschlag für eine Verordnung über die Achtung des Privatlebens und den Schutz personenbezogener Daten in der elektronischen Kommunikation und zur Aufhebung der Richtlinie 2002/58/EG) vorgestellt. Die Inhalte des Vorschlags sind daher nicht von Grund auf neu, sondern nur eine Weiterentwicklung der bereits bestehenden Regelungen. Der Abschluss des Dossiers ist für das 1. Halbjahr 2018 geplant, um ein Inkrafttreten möglichst zeitgleich mit der allgemeinen Datenschutz Grundverordnung (2016/679) zu ermöglichen.

Durch das Rechtsinstrument der unmittelbar anwendbaren Verordnung soll ein EU-weit einheitlicher Rahmen und eine Anpassung an die Datenschutz-Grund-VO erfolgen.

Ein wichtiger Aspekt des Entwurfs ist der Schutz von Inhalts- und Metadaten (Zeit und Ort): Vorgesehen sind umfassende Anonymisierungs- und Löschungsverpflichtungen, wenn nicht die ausdrückliche Zustimmung zur Datenverarbeitung vorliegt bzw. die Daten für Abrechnungszwecke benötigt werden.

Andererseits sollen den Unternehmen aber auch neue Geschäftsmöglichkeiten eröffnet werden: Wenn Nutzer die ausdrückliche Zustimmung zur Verarbeitung von Kommunikationsdaten geben (Inhalts- und/oder Metadaten), können Unternehmen damit auch neue Dienste, wie z.B. Heat Maps, anbieten. Dies könnte etwa für Verkehrsbetriebe hinsichtlich der Planung von Infrastrukturprojekten hilfreich sein.)

Vorgesehen sind auch einfachere Regeln betreffend Cookies, da die derzeitige Regelung zur Überflutung der Nutzer mit Zustimmungsanfragen führte. Nun soll klargestellt werden, dass Browser Einstellungen mit entsprechender Vorabinformation für den Nutzer für die Zustimmung oder Ablehnung ausreicht. Keine explizite Zustimmung soll jedoch nötig sein für „non-privacy intrusive Cookies“ (also nicht in die Privatsphäre eindringende Cookies), wie beispielsweise Cookies, die sich den Inhalt des Warenkorbs während des Online-Shoppings merken oder vom Anbieter selbst nur zur Zählung der website-Besucher verwendet werden.

Ein zentraler Punkt ist nach wie vor der Schutz vor unerbetenen Nachrichten – (E-Mail, SMS, automatisierte Anrufmaschinen): Für Werbeanrufe kann jeder Mitgliedstaat festlegen, ob diese grundsätzlich verboten sind (ohne Vorab-Zustimmung) oder die Möglichkeit der Nutzung einer do-not-call Liste vorgesehen wird.

Zur Vollziehung der Vertraulichkeitsregeln sollen die Datenschutzbehörden zuständig sein, die bereits nach der allgemeinen Datenschutz-Grund-VO eingerichtet wurden.

### 4.2.2.3 Problematik Telekom-Unternehmen gegenüber OTT

Seit der letzten Überprüfung der e-Datenschutz-Richtlinie im Jahr 2009 haben sich wichtige technische und wirtschaftliche Entwicklungen auf dem Markt vollzogen. Verbraucher und Unternehmen nutzen mittlerweile nicht mehr nur herkömmliche Kommunikationsdienste für Sprachtelefonie, Textnachrichten (SMS) und E-Mail, sondern zunehmend neue, funktional gleichwertige Internetdienste, die eine interpersonelle Kommunikation ermöglichen, wie z. B. VoIP-Telefonie, Sofortnachrichtenübermittlung (Instant-Messaging) und webgestützte E-Mail-Services.

#### 4.2.2.3.1 Was sind OTT Dienste?

OTT-Dienste (Over-the-Top-Dienste) sind Dienste, die Anwendungen über das Internet anbieten. Entscheidend dabei ist nicht die Art des Dienstes, sondern allein der Übertragungsweg. Zentrale Beispiele dafür sind Voice over IP Dienste wie Skype oder WhatsApp.

#### 4.2.2.3.2 Worin besteht die Problematik?

Over-the-Top-Kommunikationsdienste werden im Allgemeinen vom gegenwärtigen Rechtsrahmen der Europäischen Union für die elektronische Kommunikation, einschließlich der e-Datenschutz-Richtlinie, nicht erfasst. Folglich hat die e-Datenschutz-Richtlinie mit der technischen Entwicklung nicht Schritt gehalten, was zu einem mangelnden Schutz der über solche neuen Dienste abgewickelten Kommunikation führt.

Zur Gewährleistung eines wirksamen und EU-weit einheitlichen Schutzes der Endnutzer bei der Benutzung funktional gleichwertiger Dienste enthält der Verordnungsvorschlag die in der (sich ebenfalls im Verhandlungsstadium befindenden) Richtlinie des Europäischen Parlaments und des Rates über den europäischen Kodex für die elektronische Kommunikation vorgesehene Begriffsbestimmung für elektronische Kommunikationsdienste.

Diese Begriffsbestimmung erfasst nicht nur Internetzugangsdienste und Dienste, die ganz oder teilweise in der Übertragung von Signalen bestehen, sondern auch interpersonelle Kommunikationsdienste, die nummerngebunden oder nummernunabhängig sein können. Der Schutz der Vertraulichkeit der Kommunikation ist nach den Ausführungen der Kommission auch im Hinblick auf interpersonelle Kommunikationsdienste, die nur eine untergeordnete Nebenfunktion eines anderen Dienstes darstellen, unverzichtbar; deshalb sollten derartige Dienste, die auch eine Kommunikationsfunktion aufweisen, ebenfalls unter die Verordnung fallen.

Die Regelungen im Bereich des Datenschutzes sollen nun also auch für neue Marktteilnehmer, wie z.B. WhatsApp, FB-Messenger, Skype, gelten. Dadurch soll ein „level playing field“ im Verhältnis zu den traditionellen Telekombetreibern geschaffen werden. Der zentrale Streit über die OTT Dienste wird sich mit der Frage beschäftigen, welche der Beschränkungen für nummerngebundene Dienste auch für nicht nummerngebundene Dienste anwendbar sein sollen. Diese Diskussion beginnt eben erst, wobei die Betreiber naturgemäß einen kleineren Katalog an Beschränkungen bevorzugen, als dies im gegenwärtigen Entwurfstext der Fall ist.

### 4.2.2.4 Zeitplan

Die Europäische Kommission hat den E-Privacy Vorschlag am 10. Jänner 2017 vorgestellt und an den Rat und das Europäische Parlament zur Behandlung übermittelt.

Die Europäische Kommission wünscht sich einen Abschluss des Dossiers im 1. Halbjahr 2018, um ein in Kraft treten möglichst zeitgleich mit der allgemeinen Datenschutz-Grundverordnung (2016/679) zu ermöglichen.

Das erscheint jedoch wenig realistisch.

Die estnische Ratspräsidentschaft wird das Dossier zwar weiterbearbeiten, es ist jedoch im besten Fall davon auszugehen, dass beim Rat am 4. Dezember 2017 eine allgemeine Ausrichtung erzielt werden kann. Erst danach wird es zu Trilog Gesprächen mit dem Europäischen Parlament kommen, um die Positionen anzunähern. Eine rasche Einigung mit dem EP in erster Lesung erscheint unwahrscheinlich, wobei derzeit absolut nicht absehbar ist, wie die Positionierung des EP zu dem Dossier aussehen wird.

Realistischer erscheint ein Abschluss des Dossiers unter Österreichischer Präsidentschaft im 2. Halbjahr. 2018 oder erst im 1. Halbjahr 2019.

## 4.3 US vs. Europa

### 4.3.1 Bisherige Datenschutzbestimmungen im Vergleich

Wenn man sich die Datenschutzbestimmungen der Europäischen Union (EU) und der USA ansieht, wird schnell klar, dass es einige Unterschiede gibt. Einige dieser Unterschiede werden im Folgenden kurz erläutert.

Der erste und wahrscheinlich einer der größten Unterschiede betrifft die Regelung des Datenschutzes. In den USA gibt es hunderte sektoren- industrie- und gefahrenspezifische Datenschutzgesetze. Diese Gesetze können entweder auf Bundesebene oder auf Ebene der Bundesstaaten geregelt sein. (siehe Tabelle 1)Mit diesen spezifischen Regelungen kann es unter Umständen einfacher sein auf spezifische Fälle des Datenschutzes einzugehen (Determann, (2016)). Allerdings kann es zu Konflikten zwischen den sektorenspezifischen Regelungen auf Bundesebene und den staatenspezifischen Datenschutzgesetzen auf Bundesstaatenebene kommen. Das geschieht dadurch, dass manche Gesetze auf Bundesebene jenen auf Bundesstaatenebene vorausgehen, und manche Gesetze auf Bundesstaatenebene jenen auf Bundesebene vorausgehen. Dadurch sind Unternehmen oft im Konflikt nach welchen Gesetzen sie sich richten sollen. Es gibt also keine umfassende Regelung für den Umgang mit privaten Daten, dafür aber Regelungen für bestimmte Bereiche. Beispiele dafür wären der Children's Online Privacy Protection Act, oder der Health Insurance Portability and Accountability Act.[3]

In der EU gibt es im Vergleich zu den USA allgemeine Datenschutzbestimmungen, bei denen es schwieriger ist sie auf individuelle Ausnahmen und Einzelfälle anzuwenden, allerdings hat man ein Grundgesetz auf das man sich verlassen kann und es kommt zu keinen Konflikten verschiedener Gesetze.

[3] https://www.datenschutz-wiki.de/Datenschutz#Vereinigte_Staaten

*Tabelle 1: Ebenen der Datenschutzregelungen in den USA*

| Welche Gesetze regeln das Sammeln und Nutzen von personenbezogenen Daten in den USA? | | | |
|---|---|---|---|
| Art der Gesetze | Ebene | Information | Beispiele |
| Grundgesetze | federal level | Es gibt in den USA kein Grundgesetz, das den Datenschutz regelt. | - |
| Sektorenspezifische Gesetze | federal level | Es gibt eine Reihe an datenschutzbezogenen Gesetzen. Manche gelten für bestimmte Kategorien von Informationen, andere betreffen bestimmte Aktivitäten, die personenbezogene Daten verwenden, und wieder andere, wie Gesetze zum Konsumentenschutz, betreffen nicht per se den Datenschutz, aber wurden schon darauf angewendet. | The Federal Trade Commission Act<br>The Health Insurance Portability and Accountability Act<br>The Electronic Communications Privacy Act<br>etc. |
| Andere Gesetze und Richtlinien | industry level | Es gibt viele Richtlinien, die von unterschiedlichen Industrien als „best practices“ gelten. | self-regulatory programme for online behavioural advertising der Werbeindustrie |
| Staatenspezifische Datenschutzgesetze | state level | Es gibt viele Gesetze auf Ebene der Bundesstaaten, die das Sammeln und Nutzen personenbezogene Daten regeln. Kalifornien ist hier Vorreiter was Datenschutzbestimmungen angeht. | the California Electronic Communications Privacy Act |

Ein weiterer großer Unterschied zwischen den Datenschutzregelungen der EU und den USA betrifft die automatische Verarbeitung von personenbezogenen Daten. Unternehmen ist es in der EU generell verboten personenbezogene Daten von Personen zu verarbeiten, außer es handelt sich um eine gesetzlich festgelegte Ausnahme. In den USA wurde in den 70er Jahren ebenfalls über eine solche generelle Regelung nachgedacht, allerdings kam man zu dem Entschluss, dass ein solches Verbot hinderlich für die Entwicklung und Innovation von Informationstechnologien sein würde. Es gibt aber Einschränkungen mittels der sektorenspezifischen Gesetze und der staatenspezifischen Gesetze, die die Verarbeitung von Daten je nach Fall regeln. (Schwartz, (2008)) Auch das Recht auf freie Meinungsäußerung, das in den USA im 1. Zusatzartikel zur Verfassung der Vereinigten Staaten geregelt wird, kam bei dieser Entscheidung zur Sprache.

Der Zusatzartikel besagt:

> „*Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of*

*the people peaceably to assemble, and to petition the Government for a redress of grievances*"[4]

Eine generelle Regelung zum Datenschutz würde zu Einschränkungen führen und somit mit der Meinungsfreiheit kollidieren. (Determann, (2016))

Die Anwendung der Datenschutzgesetze ist ebenfalls ein Punkt, indem sich die EU und die USA unterscheiden. Während es in der EU eigene Datenschutzbehörden oder bestimmte Registrierungspflichten gibt, kümmern sich in den USA die Behörden, die sich um den allgemeinen Gesetzesvollzug kümmern, auch um Datenschutzangelegenheiten.

Auch wenn es um den internationalen Transfer von Daten geht, unterscheiden sich die Europäische Union und die USA. So unterstützen die USA den weltweiten Handel von Daten, die EU ist hier weitaus kritischer. (Determann, (2016)) Im Jahr 2000 versuchte man erstmals die Übermittlungsmöglichkeit personenbezogener Daten von der EU in die USA mit dem Safe-Harbor-Abkommen festzulegen. Dieses Abkommen wurde allerdings am 6. Oktober 2015 vom Europäischen Gerichtshof für ungültig erklärt. Am 12. Juli 2016 hat die Europäische Kommission, das EU-US-Datenschutzschild (eng.: EU-US-Privacy Shield) als Nachfolger des Safe-Harbor-Abkommens mit strengeren Richtlinien angenommen. US-Unternehmer haben die Möglichkeit sich in die „Privacy Shield List" vom US-Handelsministerium eintragen zu lassen, wenn sie sich zur Einhaltung der „Privacy Shield Principles" gegenüber dem US-Handelsministerium verpflichten. Dadurch ist der Datentransfer von der EU in die USA wesentlich einfacher. Eine Übertragung von Daten ist auch ohne Eintragung möglich, allerdings unterliegt diese dann weitaus strengeren Regelungen.[5]

Alles in Allem gibt es einige Unterschiede zwischen den Datenschutzbestimmungen der Europäischen Union und jenen der USA. Der Größte ist wohl, dass die EU, auch mit der neuen EU-Datenschutz-Grundverordnung, die im Mai 2018 in Kraft tritt, stark darauf abzielt ihre Regelungen für alle Mitgliedstaaten und Themenbereiche zu harmonisieren, während in den USA und ihren Bestimmungen sehr auf Individualität, Freiheit und das Individuum gesetzt wird.

### 4.3.2 Situation nach Inkrafttreten der DS-GVO

Laut Art 3 (2) finden die Bestimmungen der DSGVO auch Anwendung auf die Verarbeitung personenbezogener Daten von betroffenen Personen, die sich in der Union befinden, durch einen nicht in der Union niedergelassenen Verantwortlichen oder Auftragsverarbeiter, wenn die Datenverarbeitung im Zusammenhang damit steht

a) betroffenen Personen in der Union Waren oder Dienstleistungen anzubieten, unabhängig davon, ob von diesen betroffenen Personen eine Zahlung zu leisten ist;

b) das Verhalten betroffener Personen zu beobachten, soweit ihr Verhalten in der Union erfolgt.

Somit unterliegen auch etwa US-Anbieter von Online-Services, die europäische Kunden haben, den Bestimmungen der DS-GVO. Abgesehen von der Frage der Rechtsdurchsetzung wird somit bezüglich der Anforderungen die Datenverarbeitung ein „Level Playing Field" geschaffen, da auch bei einem einzigen europäischen Kunden die entsprechenden Prozesse zum Einholen der Einwilligung, zur Herstellung der Transparenz etc. zu implementieren sind. Generell stellt sich die Frage ob diese Anbieter dann diese Regeln für alle BenutzerInnen implementieren werden, zumal

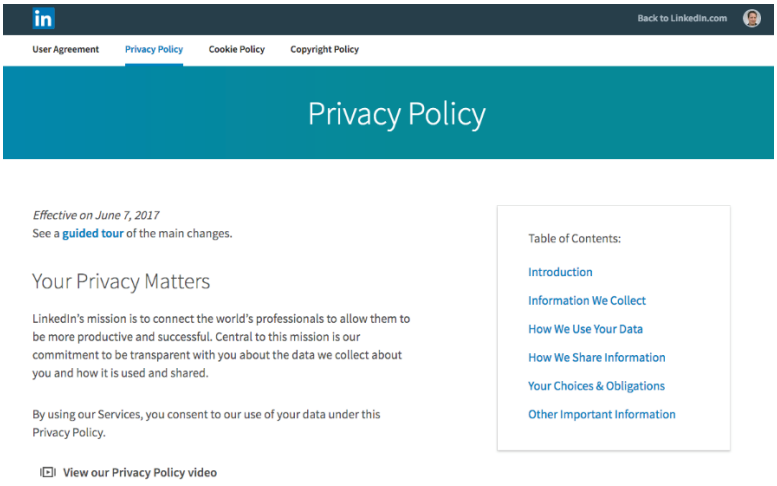
[4] First Amendment to the Constitution of the United States
[5] https://www.wko.at/service/wirtschaftsrecht-gewerberecht/EU-US-Privacy-Shield.html

eine Unterscheidung zwischen europäischen und außereuropäischen Nutzern z.B. aufgrund der Verwendung von VPN nicht einfach ist und unterschiedliche Benutzerschnittstellen problematisch sind.

Verschiedene große US-Anbieter von Online-Services, die personenbezogene Daten verwalten oder verarbeiten, reagieren bereits pro-aktiv auf die DSGVO, etwa im Sinne von geänderten Nutzungsbestimmungen. Die BenutzerInnen werden beim login zu sogenannten "privacy review" gebeten und damit verbunden aufgefordert, aktive Zustimmung zu diesen Nutzungsbestimmungen zu leisten, sowie modulare Einverständniserklärungen mittels Opt-In zu bestimmten Teilaspekten der entsprechenden Dienste zu geben. Es darf angenommen werden, dass damit sichergestellt werden soll, dass "informed consent" mit Inkrafttreten der DS-GVO in dokumentierbarer Form vorliegt.

Als Beispiel hat etwa LinkedIn im Juni 2017 seine Privacy Policy adaptiert,[6] wobei auch ein anschauliches Video zur Verfügung gestellt wird, das die Privacy Policy in verständlicher Form erklären soll.

*Abbildung 3: Die neue Privacy Policy von LinkedIn*



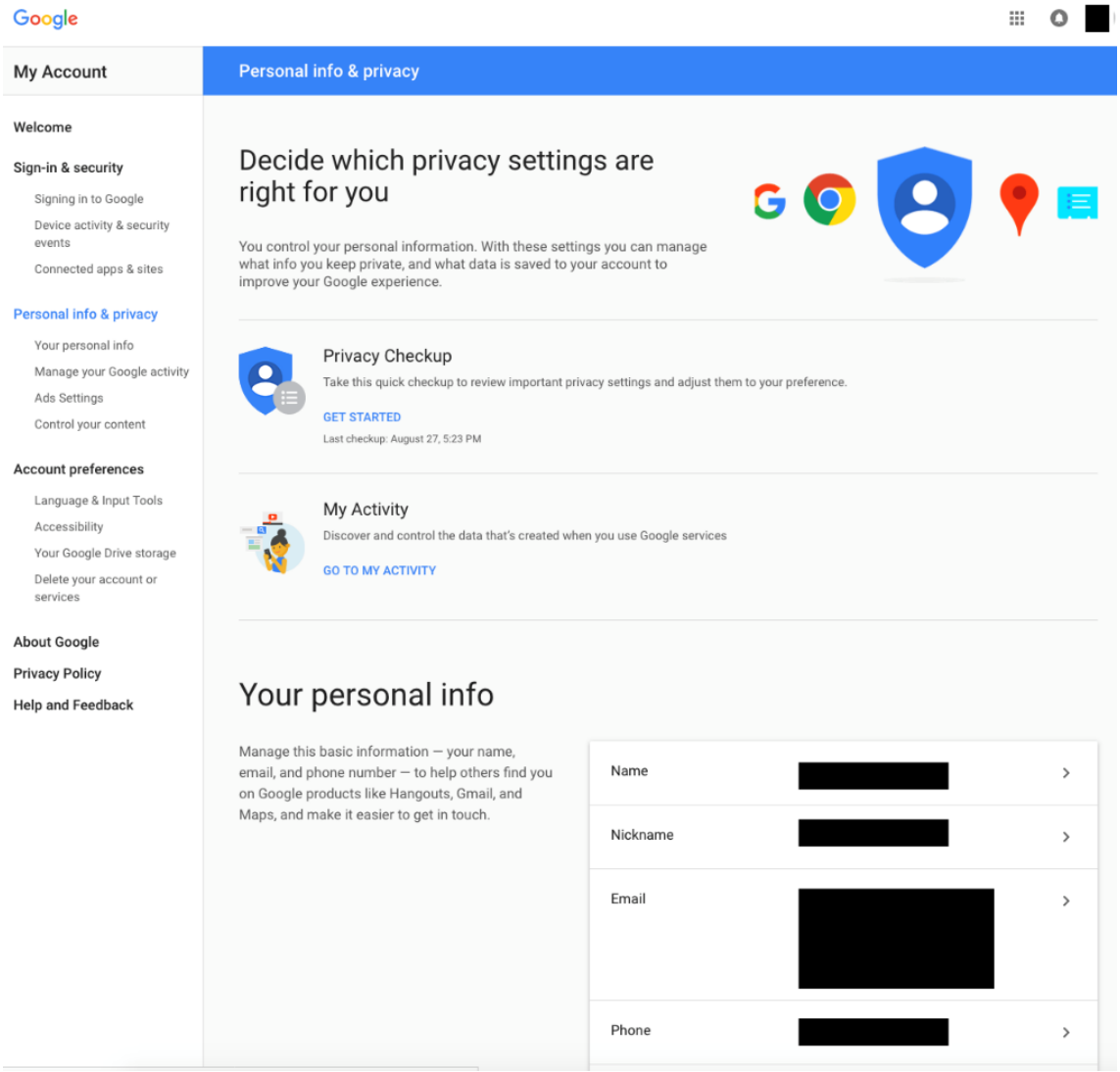

Hierbei werden den BenutzerInnen zahlreiche Auswahlmöglichkeiten geboten, die Speicherung personenbezogener Daten zu kontrollieren bzw. diese zu löschen.

Ein ähnliches Beispiel bietet Google's Dashboard. Auch hier hat Google vor kurzem seine BenutzerInnen nach dem Login aufgefordert, die Privacy-settings aktiv im Rahmen eines "Privacy Checkup" über ein intuitives User Interface zu reviewen, wobei wiederum angenommen werden kann, dass dies geschah um informierte Zustimmung zur Speicherung personenbezogener Daten einzuholen, und den BenutzerInnen diverse Auswahlmöglichkeiten geboten werden, diese Zustimmung jederzeit einzuschränken bzw. personenbezogene Daten herunterzuladen.

Es scheint also, als wären gerade die großen nicht-europäischen Service-Anbieter derentwegen die Datenschutzgrundverordnung ins Leben gerufen wurde, diejenigen die – zumindest nach außen hin - bereits am intensivsten an deren Umsetzung arbeiten. Sie verfügen allerdings auch über eine kritische Masse an BenutzerInnen, die leicht über online-Interfaces erreicht werden können wodurch die Einholung der Zustimmungserklärungen leicht automatisiert werden kann.

[6] https://www.linkedin.com/legal/privacy-policy, letzter Zugriff 9.9.201792017

*Abbildung 4: Googles "Privacy Dashboard*




Obwohl in beiden diesen Beispielen die Intention klar ersichtlich ist, die Anforderungen der DS-GVO in adäquater Art und Weise zu erfüllen, bestehen hier noch gute Chancen für österreichische und europäische Anbieter, bessere und der Intention der DS-GVO besser entsprechende Lösungen anzubieten. Bei LinkedIn wird etwa im Rahmen der Löschmöglichkeiten personenbezogener Daten auch darauf hingewiesen, dass gewisse Daten selbst nach Löschung des Accounts weiterhin gespeichert bleiben - etwa um weitere “regulatory requirements” zu erfüllen. Da US-Firmen an andere Regularien gebunden sind als europäische, besteht die Möglichkeit, dass es hier zu Konflikten kommen könnte.

# 5 Technische Analyse

## 5.1 Ausgewählte technische Aspekte des Datenschutzes

Im folgenden Kapitel beschäftigen wir uns mit einigen technischen Fragestellungen, die speziell auf Basis der rechtlichen Analyse auftauchen und entsprechend Auswirkungen auf die tatsächliche Umsetzung der EU-Datenschutzgrundverordnung (DSGV) besitzen. Dabei sollen nicht nur die Probleme, bzw. technischen Gegebenheiten, die einer weiteren rechtlichen Beantwortung durch Gesetzgeber und Judikatur bedürfen, beschrieben werden, wir geben auch Hinweise und mögliche Maßnahmen zur Umsetzung an. Neben Grundstrategien bediene wir uns dabei eines stufenweisen Ansatzes, d.h. je nach Wahl des akzeptablen Restrisikos, sowie weiterer Einschränkungen durch andere Regularien, wie bspw. im Finanzwesen, werden entsprechend strikte rechtliche Auslegungen technisch umgesetzt.

Als ausgewählte Fragestellungen wurden im Rahmen von Workshops mit österreichischen Unternehmen und Behörden vordringlich die folgenden Themen erachtet:

- Löschung von Daten und Informationen, speziell in Hinblick auf das in der DSGV verankerte „Recht auf Vergessenwerden“.
- Anonymisierungsstrategien und ihre Eigenschaften, Einschränkungen und impliziten Annahmen, die in Hinblick auf ihre Gültigkeit im jeweiligen Kontext hinterfragt werden müssen.
- Das Problem des Antagonismus zwischen dem „Recht auf Vergessenwerden“ und Forderungen zur Transparenz in der Datenverarbeitung. Letztes ist nicht nur Gegenstand zahlreicher branchenspezifischer Regularien (Basel 2, SOX, HIPAA), sondern erwächst auch aus der DSGV selbst. Des Weiteren entsteht ein beträchtlicher Aufwand, bzw. bestehen verschiedene technische Möglichkeiten (im Sinne von Architekturen zur Speicherung von Transparenz-Informationen, standardisierten Austauschformaten, etc.) für zur Realisierung der Einhaltung der in der DSGV geforderten Transparenzanforderungen und damit verbundener Speicherung entsprechender Aufzeichnungen zur Verwendung personenbezogener Daten.
- Die Effekte der Löschung von Daten, aber auch der Anonymisierung in Hinblick auf Big Data-Anwendungen, speziell in Hinblick auf die Effekte in Richtung Verfälschung. Dieser Aspekt ist ganz wesentlich für die Beurteilung des Effekts dieser Regularien auf die Nutzbarkeit von Daten zur Generierung innovativer Services und Produkte.
- Zu guter Letzt diskutieren wir technische Aspekte der Realisierung der Einholung expliziter, unmissverständlicher, informierter Einwilligung („informed consent“) zur zweckgebundenen Verarbeitung personenbezogener Daten, wie in der DSGV gefordert.

Im Rahmen dieses Kapitels werden neben Lösungsmöglichkeiten auch derzeit technisch nicht gelöste Probleme dargelegt, bei denen noch entsprechender Forschungsbedarf besteht. Grundsätzlich muss allerdings gesagt werden, dass viele Probleme derzeit auch in der starken Formulierung der Regulierungstexte begründet liegen, die in dieser Absolutheit grundsätzlich nur schwer bis unmöglich erfüllbar sind. Hier gehen wir davon aus, dass sich die Judikatur der Realität anpassen wird, wie dies auch schon im Rahmen der Verschlüsselung gemacht wurde und die

Nutzung des State-of-the-Art in den Bereichen Algorithmik und existierenden Werkzeugen und Software-Lösungen als ausreichend angesehen wird um vor etwaiger gerichtlicher Verfolgung geschützt zu sein.

## 5.2 Löschung von Daten und Informationen

Das Recht auf Vergessenwerden, sowie die EU DSGV (General Data Protection Regulation), zeichnen sich dadurch aus, dass sie es, in einem gewissen Rahmen, ermöglichen, sensible und personenbezogene Daten aus datenverarbeitenden Umgebungen und Anwendungen zu löschen. So simpel diese Vorschrift erscheinen mag, eröffnet sie ein extrem spannendes, herausforderndes, aber auch (rechtlich) unklares Feld. Dies liegt vor allem darin begründet, dass der Begriff des Löschens nicht ausreichend definiert wurde.

Grundsätzlich wird der Terminus „Löschen“ in Datenanwendungen ganz unterschiedlich verwendet, mit weitreichenden impliziten Folgen. In vielen Datenanwendungen und Produkten wird dabei keine endgültige Vernichtung („physikalisches Löschen“) der Daten, wie durch das Wort suggeriert, gemeint, sondern lediglich eine Entfernung aus der Informationsverarbeitung („logisches Löschen“). Die betrifft nicht nur Datenbanken und andere datenfokussierte Software, sondern im Endeffekt auch auf die meisten Betriebssysteme und deren Dateisysteme zu – diese werden ebenfalls typischerweise nicht überschrieben, es wird lediglich der Speicherplatz im internen Inhaltsverzeichnis (TOC – table of content) als frei markiert und kann somit bei einem Schreibvorgang wieder genutzt werden. Forensische Tools, die solcherart „gelöschte“ Dateien wiederherstellen können sind jedoch schon seit vielen Jahrzehnten verbreitet. Der Erfolg einer Wiederherstellung hängt im Wesentlichen davon ab, ob die Blöcke schon wieder genutzt wurden. Dies wiederum hängt im Wesentlichen von der seit der Löschung vergangenen Zeit und der Nutzungsintensität im Sinn der Speicherung neuer Daten auf der Disk ab.

Um diese Wiederherstellung unmöglich zu machen, wurde in der Vergangenheit ein Set an Möglichkeiten der „endgültigen“ Löschung entwickelt, von denen das Überschreiben der Speicherbereiche mit Zufallswerten die populärste ist. Für diese Form gibt es auch eine umfangreiche Sammlung an Tools. Obwohl es in diesem Bereich durchaus noch Diskussionen über die richtige Form des Überschreibens gibt (random, patterns, mehrfach), da nach einmaligem Überschreiben mit fixen Mustern in akademischen Labors noch ursprüngliche Daten wiederhergestellt werden konnten, gilt diese Methode in der praktischen Anwendung als hinreichend sicher um von einer endgültigen Löschung der Dateien ausgehen zu dürfen. Dies kann allerdings schon bei Einsatz von Flash-Speichern (bspw. USB-Sticks) hinterfragt werden, da durch Gegenmaßnahmen gegen das Ware-Leveling (das ist die Abnutzung der Speicherzellen bei Flash-Speichern durch Beschreiben) die Speicherzellen nicht mehr direkt beschrieben werden können, sondern diese vielmehr möglichst gleichmäßig genutzt und teilweise für weitere Bearbeitung gesperrt werden – die gelöschten Daten werden daher in solchen Blöcken nicht weiter überschrieben. Ware-Leveling beschreibt dabei das Problem von Flash-Speichern, dass die Datenzellen nicht beliebig oft genutzt werden können, sondern sich nach einigen tausend Schreibzyklen abnutzen und dann unbrauchbar werden. USB-Sticks trachten daher danach, die Speicherzellen möglichst gleichmäßig abzunutzen. Dennoch kann man auch in diesen Systemen technisch gesehen von einer grundsätzlichen Möglichkeit der Löschung ausgehen.

Allerdings hat es sich in extrem kritischen Bereichen durchaus die physische Zerstörung von Datenträgern eingebürgert. Dies ist speziell dann sinnvoll, wenn eine lange vorgehaltene große Datenmenge als Gesamtheit endgültig vernichtet werden muss. Dieser Zugang ist allerdings für die Umsetzung eines Rechts auf Vergessenwerden, bei dem es lediglich um die Löschung

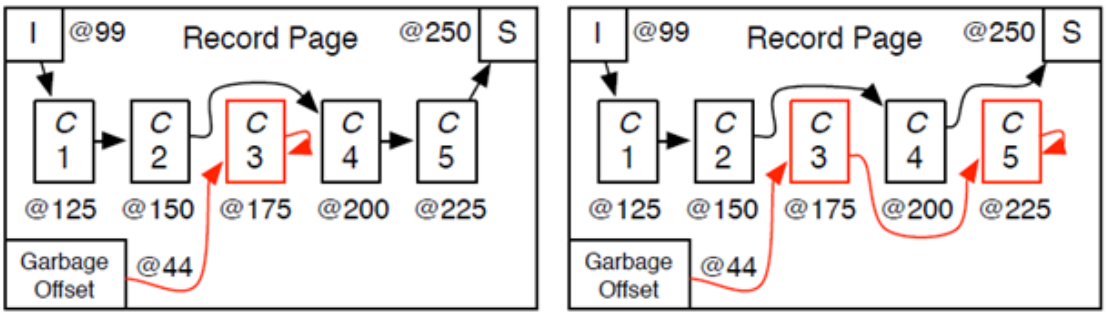
einzelner oder einer kleinen Menge an sensiblen Informationen geht, wirtschaftlich nicht vertretbar und in hochverfügbaren Systemen auch technisch nicht realisierbar.

### 5.2.1 Datenlöschung in komplexen Systemen

Wesentlich komplexer sind Systeme zu handhaben, in denen Informationen nicht auf reiner Dateiebene verwaltet werden können. Besonders sind dabei für den Bereich Big Data Datenbanken hervorzuheben, da diese vordringlich der strukturierten Speicherung großer Mengen an Daten dienen und als gutes Beispiel für eine komplexe Softwareapplikation dienen können: Aus der logischen Anwendersicht gestaltet sich die Handhabung der Löschung in Datenbanken denkbar einfach: Daten werden in Tabellen gespeichert und können mittels einfacher Befehle (DELETE) aus diesen gelöscht werden. Dabei wird allerdings nicht beachtet, dass die Löschung aus Datenbanken (aus Gründen der Performance) faktisch nie ausgeführt wird, lediglich wird das Inhaltsverzeichnis der Tabelle (der Index) so verändert, dass es die Datenzelle nicht mehr enthält, die Daten werden jedoch nicht überschrieben. Dies geschieht grundsätzlich analog zu den Mechanismen in vielen Dateisystemen, allerdings mit den folgenden generellen Unterschieden am Beispiel von MySQL:

- Da es auch notwendig sein kann Löschungen rückgängig zu machen, muss der gelöschte Inhalt in entsprechenden Mechanismen vorgehalten werden.
- Die gelöschten Zellen werden in der Datenbank in einem eigenen Index verwaltet, der sogenannten Garbage Collection, und somit nicht nur bezüglich ihres Inhalts, sondern auch der Lösch-Timeline analysiert werden (siehe auch Abbildung 3).
- In Versuchen konnten wir feststellen, dass die gelöschten Speicherzellen noch relativ lange nicht genutzt werden. Dies dürfte darauf zurückzuführen sein, dass es für die Datenbank performanter ist, neuen Platz in einem großen freien Speicherbereich (bspw. am Ende des DB-Files) zu allokieren, als einen passenden Bereich aus der Garbage Collection zu suchen.

*Abbildung 5: Löschen in Datenbanken*




(a) One deleted record (b) Two deleted record

Zusätzlich speichert eine Datenbankanwendung die Informationen nicht nur in den Tabellen selbst, sie speichert auch Informationen zu Veränderungen in Log-Dateien, um die Veränderungen der Datenbank nachvollziehbar zu machen. Viele dieser Log-Informationen dienen dem Datenbankadministrator auch zur Überwachung der Operationen einer Datenbank und stellen ein wesentliches Merkmal zur Sicherstellung der Integrität und der Transparenz dar. Speziell interne Mechanismen enthalten ebenfalls Informationen, um grundlegende Eigenschaften wir Crash-Recovery und Rollback umsetzen zu können. Die Informationen in

diesen Mechanismen sind dabei durchaus in der Lage, längere Zeit zu überdauern, speziell wenn die Zahl der schreibenden Operationen auf einer Datenbank gering ist.

In Hinblick auf die Löschung von Daten aus komplexen Systemen gibt es daher grundsätzlich zwei Möglichkeiten:

1. Die Vernichtung des Gesamtsystems, d.h. Löschung aller Speichermedien, sowie Backups, Replikationen und dergleichen, mit anschließender Überschreibung der Speicherbereiche oder physikalischer Vernichtung. Eine etwas praktischere Version dieser Technik wird im nächsten Kapitel besprochen.

2. Tiefgehende Analyse der Applikationen und entsprechende manuelle Löschung. Dabei muss natürlich ein entsprechendes technisches Potenzial zur Wiederherstellung angenommen und entsprechend begründet werden. Dies ist vor allem dann unumgänglich, wenn, wie im Rahmen des Rechts auf Vergessenwerden vorgeschrieben, Einzeldaten gelöscht werden müssen.

Möglichkeiten zur Umsetzung der ersten Option finden sich im folgenden Unterkapitel, hier soll vor allem die Problematik der Löschung individueller Datensätze diskutiert werden. Ein sinnvolles Vorgehen umfasst dabei die Erstellung eines aus der Security bekannten Angreifer-Modells, d.h. die Abschätzung der Fähigkeiten eines eventuellen Angreifers (in unserem Fall nicht unbedingt klassisch als Angreifer zu verstehen, sondern vielmehr als eine Person und oder Institution, die danach trachtet gelöschte Informationen wiederherzustellen). Daraus und aus der bekannten Systemarchitektur kann eine Reihe von Angriffsvektoren hergeleitet werden und damit die Angriffsoberfläche spezifiziert werden. Auf Basis dieser Analysen wird anschließend eruiert, ob die Mittel zu einer Begegnung diese Angriffsvektoren zur Verfügung stehen und realistisch (d.h. auch kommerziell) sinnvoll umsetzbar sind. Ist dies nicht der Fall, müssen etwaige Änderungen am System durchgeführt werden, bspw. durch die Wahl von Alternativprodukten, oder aber der Einstellung der Verarbeitung.

Um NutzerInnen eine mögliche Auswahl bezüglich des Sicherheitsniveaus zu geben, werden die folgenden grundlegenden Angreifer-Modelle gewählt. Durch die notwendige Verallgemeinerung können natürlich die Rollen nicht zu genau beschrieben werden, wir werden daher im Anschluss noch das wichtige Beispiel einer relationalen Datenbank genauer beschreiben.

- ***Externer passiver Beobachter an der Kommunikationsleitung***: Dieser Angreifer kann nicht direkt auf die Datenanwendung zugreifen, sondern lediglich die Kommunikationskanäle von außen beobachten. Dies betrifft unter Umständen die Anlieferung von Daten und die Extraktion von Analyseergebnissen. Diesem Angreifer kann dadurch begegnet werden, dass die Kommunikationskanäle entsprechend gesichert werden, oder aber die Verarbeitungsumgebung gänzlich von extern erreichbaren Netzwerken entkoppelt wird. Auch hierbei treten natürlich klassische IT-Security-Probleme auf (siehe auch Stuxnet, diese Malware wurde entsprechend angepasst um eine an sich komplett isolierte Umgebung zu infiltrieren, wobei typischerweise u.a. infizierte USB-Sticks genutzt werden). Der passive Beobachter ist in Hinblick auf die Löschung ein Spezialfall des passiven Nutzers, der die zusätzliche Einschränkung besitzt, dass er selbst keinerlei Abfragen durchführen kann, sondern lediglich passiv beobachtet.

- ***Normaler passiver Nutzer der Datenanwendung***: Ein Nutzer, der sich lediglich die Ergebnisse einer Datenanwendung, bspw. die Ergebnisse eines Data-Cubes ansehen

darf. Dies sind bspw. Billing- oder Controlling-Verantwortliche. Sie können lediglich die Ergebnisse sehen, die Rohdaten sind bis zu einer gewissen Aggregationsstufe nicht weiter auflösbar. Die Gefahr solcher Angreifer besteht, neben Angriffen auf das Berechtigungssystem im Rahmen von Privilege Escalation, hauptsächlich darin, dass durch geschickte Abfragen unterschiedlich aggregierter Datenwerte über einen gewissen Zeitraum hinweg auch auf gelöschte Daten rückgeschlossen werden kann. Dies kann sogar dazu führen, dass die Daten erst durch die Löschung selbst für Nutzer dieser Klasse sichtbar werden. Beispiel: Ein Angestellter im Projektcontrolling sieht die Liste der Projektmitarbeiter, sowie deren durchschnittliche Kosten pro Stunde. Wird der Datensatz des Mitarbeiters gelöscht, so verändern sich die durchschnittlichen Kosten, es kann daher u.U. direkt auf das Gehalt des Mitarbeiters rückgeschlossen werden.

- ***Entwickler in der Datenanwendung***: Viele Datenanwendungen erlauben Script-gesteuerte Automatisierung von Tasks wie Abfragen, Optimierung, aber auch Anonymisierung, Daten-Handling und dergleichen. Zusätzlich können Entwickler oftmals auch direkt auf Daten zugreifen und anlegen, oder aber sehr schnell und automatisiert eine große Menge an Abfragen zu den richtigen Zeitpunkten starten. Je nach Mächtigkeit der Verarbeitungssprache in der jeweiligen datengetriebenen Anwendung ergibt sich mehr oder weniger Potential für diesen Angreifer.

- ***Administrator/Superuser in der Datenanwendung***: Administratoren besitzen in vielen Datenanwendungen umfangreiche Zugriffsrechte auf interne Logfiles, die beispielsweise der Transparenz der Verarbeitung dienen. Je nach Anwendung besitzen sie auch wesentlich größere Rechte, bspw. in Hinblick auf Audit Trails und dergleichen, in vielen Datenbankanwendungen können sie sogar Einträge in Logfiles unwiederbringlich und unauffällig löschen. Die Fähigkeiten dieser Nutzer sind extrem von der entsprechenden Datenanwendung abhängig, diese Gruppe ist aber insofern sehr interessant, als dass ihr in vielen datenbankbasierten Realsystemen sehr großes Vertrauen entgegengebracht wird. Aus diesem Grund fokussieren wir uns auch im Bereich der Datenbank-Antiforensik im Anschluss sehr stark auf diese Angreiferklasse.

- ***Administrator am Betriebssystem mit Eingriffsrechten auf die Datenanwendung***: Je nach Setup der Datenanwendung besitzen Administratoren am Betriebssystem sehr große Möglichkeiten. In Hinblick auf ihre Unterscheidung zu Datenbankadministratoren kann es durchaus dazu kommen, dass OS-Administratoren in der Lage sein können, bspw. Änderungen am Source-Code der Datenanwendung vorzunehmen und diese neu zu deployen. Eine derart veränderte Datenbank könnte bspw. Löschungen mitprotokollieren und weiterleiten. Hierbei hilft es, ein Mehr-Augen-Prinzip mehrerer Administratoren mit eng eingegrenzten Aufgabengebieten zu verfolgen, die sich gegenseitig kontrollieren, sowie die Hashwerte der Executables und anderer wichtiger Dateien der Datenanwendung regelmäßig zu kontrollieren.

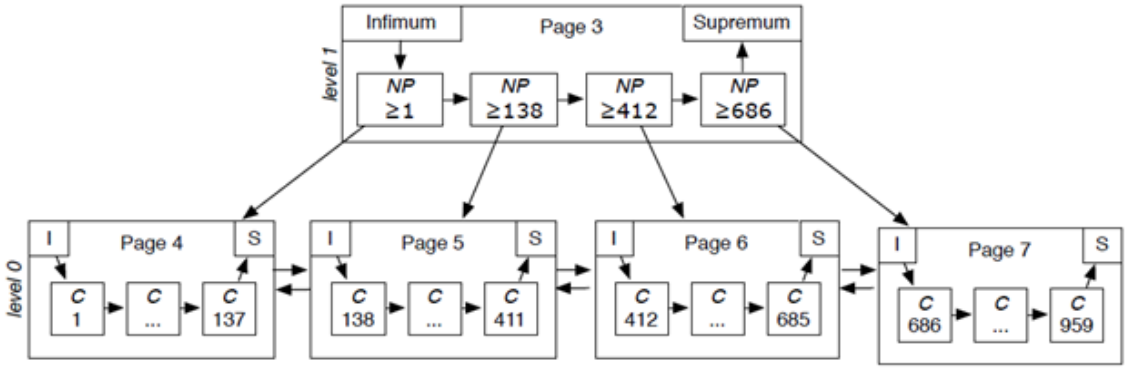
Da nahezu jede Datenanwendung, die der Verarbeitung großer Datenmengen dient, eine Form von Datenbank im Hintergrund besitzt, möchten wir hier am Beispiel von MySQL, spezifischer gesagt MySQL mit der Default-Storage-Engine InnoDB, genauer auf die Problematik des Löschens eingehen und die verschiedenen Abstufungen der Informationsgewinnung zu gelöschten Daten zeigen. Wir nehmen dazu ein Angreifer-Modell an, das über Administratorrechte auf der Datenbank und zumindest Leserechte auf die am Betriebssystem liegenden Datenbankdateien verfügt. Damit werden auch gleich mögliche Angriffe normaler

Nutzer mit abgedeckt, da diese immer über weniger Rechte als die Datenbankadministratoren verfügen.

Das erste Problem liegt an dem Verfahren der Löschung selbst. Wie schon in Abbildung 3 dargestellt, werden gelöschte Daten in Datenbanken nicht überschrieben. Genauer gesagt, sind die Daten in InnoDB (und den meisten anderen kommerziellen Datenbanken) in Form eines sogenannten $B^+$-Baums strukturiert gespeichert (Frühwirt et. al. (2015)). Ein B+-Baum ist ein balancierter Baum, der eine sehr schnelle Suche nach einem durch einen Schlüssel (dem Primärschlüssel, oder Primary Key) spezifizierten Datensatz erlaubt (siehe Abbildung 4). Jeder Tabelle in InnoDB besitzt so einen Primärschlüssel, anhand dessen auch das gesamte Speicherlayout der Tabelle in Baumform strukturiert wird.

Wird ein Datensatz gelöscht, so wird der Speicherbereich in dem die entsprechenden Daten liegen, wie in Abbildung 3 dargestellt, aus dem Inhaltsverzeichnis der Baumstruktur herausgenommen, d.h. die Nachbardatensätze lassen den Speicherbereich quasi aus und als frei markiert. Diese Markierung erfolgt dadurch, dass der Speicherbereich in die am Garbage-Offset beginnende Kette eingegliedert wird. Es erfolgt jedoch keinerlei Löschung im Sinn des Überschreibens des Speicherbereiches mit Zufallspatterns oder ähnlichen Maßnahmen. Zusätzlich kann es sehr lange dauern, bis die Datenbank einen spezifischen Speicherbereich wieder nutzt, die Gründe dafür dürfen in der Performance zu suchen sein.

*Abbildung 6: Baumstruktur des Index in InnoDB*



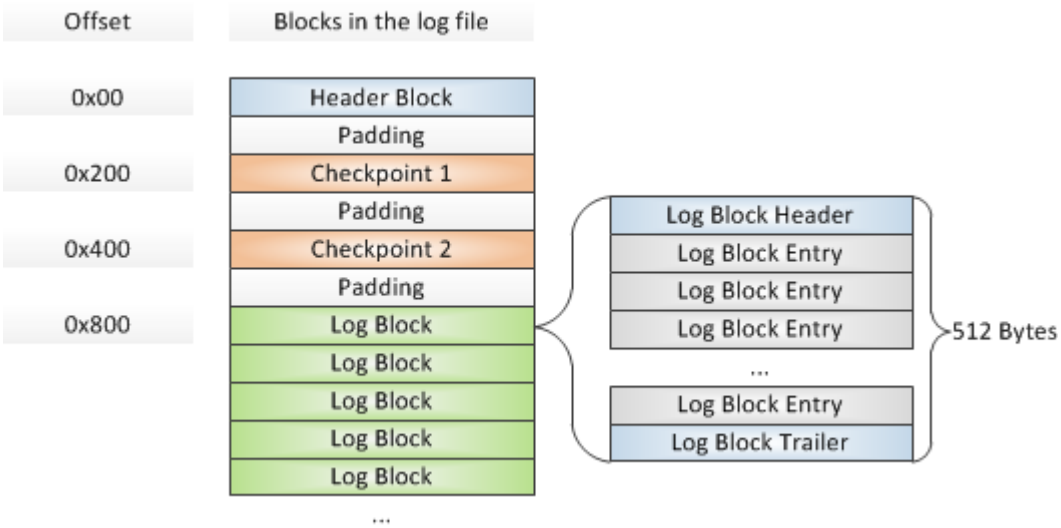

Kann ein Angreifer direkt auf den $B^+$-Baum einer Tabelle zugreifen, so ist er auch in der Lage die gelöschten Speicherbereiche zu lesen, durch die Verknüpfung mit dem Garbage-Offset kann sogar eine (bruchstückhafte) Zeitlinie der Löschung hergestellt werden. Dieser Zugriff kann direkt über das Dateisystem geschehen (Fruehwirt et. al. (2015)): Datenbanken müssen ihre Daten letztendlich ebenfalls am Dateisystem speichern, dies geschieht in den sog. Database-Files (je nach Produkt unterschiedlich). Dies sind Dateien, in denen die Baumstruktur, sowie eine Vielzahl an weiteren Informationen sehr stark strukturiert abgespeichert werden. Zusätzlich werden oftmals noch weitere Dateien zur Speicherung der logischen Tabellenstruktur, von Skripten und Prozeduren, sowie Abfragen (bspw. Views) angelegt. Kenntnisse über die interne Struktur dieser Datenbankdateien erlaubt es daher, die als gelöscht markierten Speicherbereiche anzusteuern und die darin enthaltenen Informationen zu lesen. Verschlüsselung der Datenbankdateien könnte dagegen Abhilfe schaffen und wird auch von einzelnen Herstellern propagiert (siehe bspw. ORACLE TDE[7], allerdings besitzen diese Ansätze Einfluss auf die Performance der Datenbank speziell im Fall von Big Data. Zusätzlich müssen speziell in industriellen Umgebungen noch sehr lange Legacy-Systeme in Betracht gezogen werden, die eine derartige Verarbeitung nicht

[7] http://www.oracle.com/technetwork/database/options/advanced-security/index-099011.html

unterstützen. Speziell gegen Angreifer mit Administratorrechten sind die derzeit bekannten Ansätze auch nicht hinreichend, da diese die verschlüsselten Daten in der Datenbank wiederherstellen können und dann ganz rechtmäßig auf die Entschlüsselungsroutinen zugreifen können. Eine weitere technische Lösung des Problems würde darin bestehe, die Datenbank in regelmäßigen Abständen zu reorganisieren, d.h. die Datendefragmentierung, die durch die gelöschten Datensätze entsteht, zu reduzieren. Dies kann in üblichen Systemen mit – systeminternen Programmen gelöst werden, allerdings ist die Prozedur relativ aufwändig. Für eine wirklich 100%-ige Absicherung müsste nach der Reorganisation der gesamte als frei gekennzeichnete Speicherbereich gescreent und ggf. überschrieben werden.

Ein weiteres Datenleck bezüglich gelöschter Datensätze entsteht durch die Forderung an eine Datenbank, ACID-Compliant zu sein[8], spezifischer, die Forderung, Datenmanipulation als atomare Aktionen durchzuführen (Alles-oder-Nichts-Prinzip). Dies ist essentiell, damit sich die Datenbank, speziell im Fall eines Crashs, immer in einem definierten Zustand befindet. Zusätzlich erlauben es derartige Mechanismen, Rollbacks, also das Rückgängigmachen von Aktionen, anzubieten. In InnoDB werden zu diesem Zweck alle datenverändernden Operationen (also keine Abfragen) im sog. Transaction-Log gespeichert (Fruehwirt et. al., (2012)). Dabei ist der Begriff „Log“ etwas irreführend, handelt es sich doch um einen rein datenbankinternen Mechanismus auf den kein Administrator Zugriff hat (weder lesen noch schreiben). Unter InnoDB ist dieser Mechanismus in der Form zweier Dateien gelöst, die die wesentlichen Informationen, sowie Absicherungsmechanismen gegen irrtümliche Fehler im Fall von Crashes besitzen (der Crash könnte ja auch während des Schreibens des Logs passieren, daher werden Redundanz und Hash-Values zur Absicherung der Integrität der Logdateien selbst angewandt). Die in den Dateien enthaltenen Informationen werden nicht nach einem erfolgreichen Commit gelöscht, da dies der Performance abträglich wäre, sie werden vielmehr wie Ringspeicher genutzt: Die Dateien besitzen eine fixe Größe, ist diese erreicht, wird der Zeiger wieder auf den Anfang der Datei gesetzt und dort weitergeschrieben (siehe auch Abbildung 5 zur Struktur der Logfiles).

*Abbildung 77 – Struktur der Transaction-Logs*




Da das Transaction-Log ein Rollback ermöglicht, sind nicht nur Informationen zu Dateneinfügungen und Löschungen enthalten, sondern auch die Ursprungswerte in Updates, d.h. es können durch das Transaction-Log auch mit Boardmitteln überschriebene Daten

[8] Atomarität, Konsistenz, Isolation und Dauerhaftigkeit – Anforderungen, um zu garantieren, dass die Informationen in einer Datenbank zuverlässig und persistent gespeichert sind.

wiederhergestellt werden (Fruehwirt et. al., (2013)) (allerdings nicht, falls diese im Zuge einer Reorganisation überschrieben wurden, hier geht es um reine UPDATE-Anweisungen), sowie auch die Operationshistorie. Allerdings kann dies alles nur bis zum Zeitpunkt der ältesten nicht-überschriebenen Operation geschehen, d.h. das Log „vergisst“ nach und nach die gespeicherten Operationen. Die Rate dieses Vergessens richtet sich dabei vornehmlich nach der Zahl der durchgeführten Operationen pro Zeiteinheit, sowie der Größe des Logs, hier können bspw. auf Basis einer Analyse der Datenbewegungen Maßnahmen getroffen werden, damit keine Informationen zu Daten extrahiert werden können, die vor mehr als bspw. einem Monat gelöscht wurden. Allerdings ist sehr stark vom direkten Löschen von Einträgen aus den Logfiles abzuraten, da dies die Integrität der Logfiles und damit die Integrität der gesamten Datenbank zerstört. Modifikationen direkt an den Logfiles benötigen extremes Wissen und Fingerspitzengefühl und selbst dann kann eine Zerstörung nicht ausgeschlossen werden (bspw. wenn während der Manipulation lesend auf das Log zugegriffen wird - ein Vorgang, der durchaus nicht unwahrscheinlich ist).

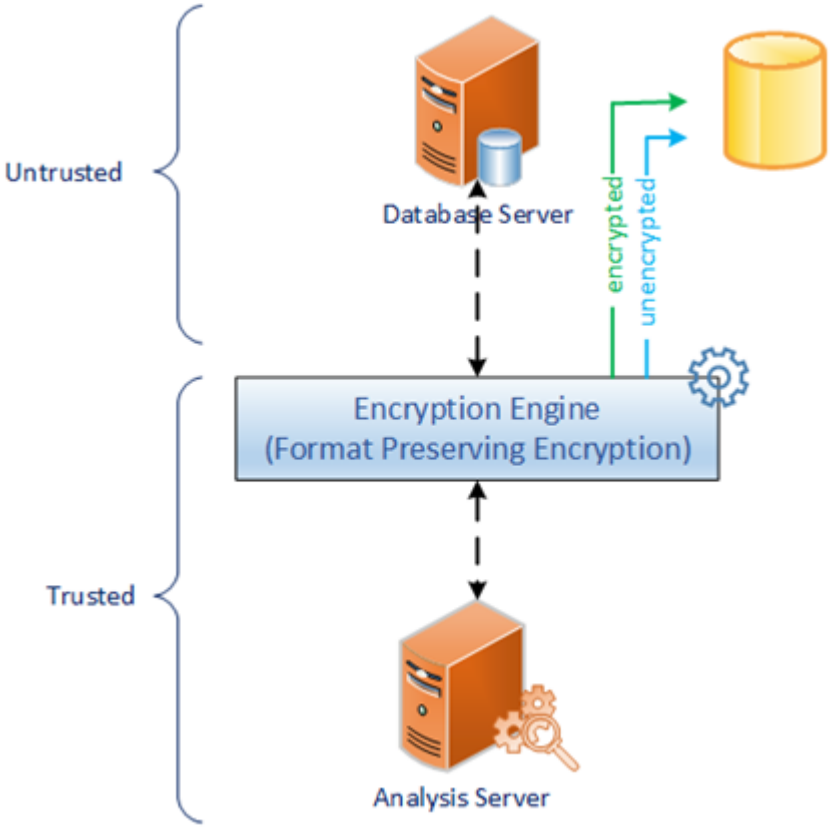
Datenbanken besitzen noch einige ähnliche Mechanismen, sehr typisch ist auch das Replication-Log. Wie auch bei Transaction-Log ist der Begriff „Log“ irreführend, handelt es sich auch hierbei um einen internen Mechanismus, diesmal allerdings nicht für Crash-Recovery und Rollback, sondern um den Mechanismus der Datenbankreplikation anzubieten. Dies bedeutet die Erstellung exakte Kopien der Datenbank, unter Umständen an anderen Orten, zum Zweck der Ausfallsicherheit. Das Datenbankreplikationslog stellt dabei sicher, dass eine Operation atomar (d.h. entweder ganz oder gar nicht) auf die replizierten Datenbanken gespiegelt wird, d.h. eine Operation wird auf der Originaldatenbank erst dann als vollzogen angesehen, wenn sie auch auf allen replizierten Entitäten vollzogen wurde. Wie auch im Fall des Transaction-Logs kann auf das Replication-Log lesend zugegriffen und entsprechend alte Meldungen zur Wiederherstellung gelöschter Informationen genutzt werden (Fruehwirt et. al., (2014)).

Zum Abschluss wollen wir noch eine sehr abstrakte Methode der Wiederherstellung gelöschter Daten vorstellen, die speziell die Problematik des Hantierens mit einem absoluten Löschbegriff verdeutlicht, auch wenn sie praktisch gesehen wenig Relevanz besitzt. Wie bereits erwähnt werden die Daten in den Datenbankdateien strukturiert in Baumform gespeichert. Allerdings ist der Baum für eine spezifische Datenmenge nicht eindeutig konstruierbar, d.h. es gibt verschiedene gültige Strukturen für die gleiche Datenmenge (siehe Abbildung 6). Wir konnten feststellen (Kieseberg et. al., (2011)), dass im Fall einer strikt monotonen Einfügereihenfolge eine bestimmte Form des Baums erstellt werden muss, d.h. ist diese Form verletzt, so kann darauf geschlossen werden, dass ein bestimmter Datenbankindex gelöscht, oder gelöscht und hinreichend später erst wieder eingefügt wurde. Derzeit ist noch nicht klar, wie sich die Strukturen im Fall anderer Operationen, speziell auch Reorganisationen, verhält, allerdings stellen die derzeit extrahierbaren Informationen typischerweise keine Gefahr der Extraktion sensitiver Informationen dar, der Ansatz ist hochtheoretisch und daher in unseren Augen derzeit vernachlässigbar.

*Abbildung 88 – Eine Menge, zwei Bäume.*

In Hinblick auf die Verwendung spezieller verschlüsselter Datenbanken, seien Ansätze dargelegt wie sie bspw. im data-centric Security Approach von HP[9] Anwendung finden (siehe Abbildung 7).

*Abbildung 99 – Schema der Nutzung einer verschlüsselten Datenbank im data-centric approach*




Dabei werden die Daten mit einem „Format Preserving Encryption“ (FPE) Schema verschlüsselt in der Datenbank abgelegt (Rogaway, (2010)), wobei die Möglichkeit besteht, einzelne Spalten unverschlüsselt abzulegen. FPE erlaubt dabei den Erhalt der Struktur der Information, d.h. bspw. die Abbildung von 10-stellige SVNs in andere 10-stellige Zahlen. Da die Verschlüsselung eindeutig ist könne sogar einfache Operationen wie Zählen des Vorkommens eines bestimmten Merkmals durchgeführt werden, allerdings handelt es sich hierbei nicht um Methoden der „Funktionalen Verschlüsselung“, d.h. mit den verschlüsselten Daten kann nicht im herkömmlichen Sinn gerechnet werden (es existiert kein Isomorphismus, der eine Berechnung auf den verschlüsselten Daten erlaubt). Diese Berechnungen müssen auf dem Analysis Server durchgeführt werden und davor in einer vertrauenswürdigen „En-/De-cryption Engine“ entschlüsselt werden. Sinnvoll eingesetzt kann diese Methode allerdings helfen, Daten zumindest zur Speicherung in der Speicherdatenbank zu schützen. Gleichzeitig muss allerdings darauf hingewiesen werden, dass viele FPE-Algorithmen derzeit noch nicht die gleiche Aufmerksamkeit erfahren haben wie traditionelle, was immer ein gewisses Sicherheitsrisiko darstellt.

### 5.2.2 Datenlöschung in virtualisierten Umgebungen

Virtualisierte Umgebungen werden immer öfter eingesetzt um (a) moderne Paradigmen wie Cloud-Computing und Shared Infrastructure nutzen zu können und (b) um einzelne Anwendungsbereiche stark voneinander abzugrenzen. Speziell der letztere Fall kann sehr

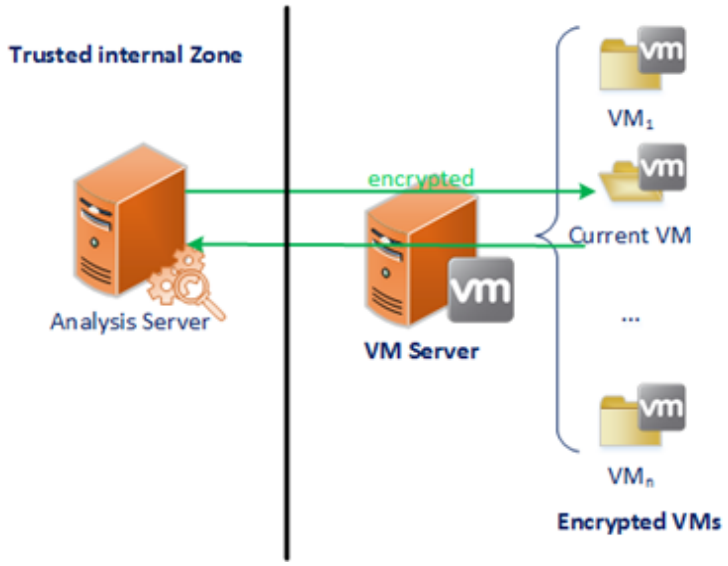
[9] https://4b0e0ccff07a2960f53e-707fda739cd414d8753e03d02c531a72.ssl.cf5.rackcdn.com/wp-content/uploads/2014/12/Voltage_WP_SecureData_Streamlining_InformationProtection_DataCentricSecurityApproach.pdf?v=20

sinnvoll sein um ein gutes Löschniveau zu erreichen. Dabei sind die folgenden Grundvoraussetzungen zu beachten:

- Die virtuelle Maschine selbst ist mit einer sicheren Verschlüsselung ausgestattet und der VM-Server ist vertrauenswürdig, d.h. bei der Ausführung der VM werden keinerlei States vom ausführenden Server gespeichert.
- Die Kommunikation mit der VM ist gut geschützt nach dem kryptographischen State-of-the-Art.
- Die Daten werden als Ganzes vernichtet, d.h. es müssen nicht Einzeldatensätze gelöscht werden, sondern die Daten werden auf die VM gespielt, dort verarbeitet und dann nicht mehr benötigt.
- Zusätzlich laufen auch alle Softwarepakete, die auf die Daten zugreifen innerhalb der verschlüsselten VM – es ist ein häufiger Fehler, dass die Daten zwar auf einer verschlüsselten Partition oder sogar auf einer verschlüsselten virtuellen Maschine gespeichert werden, die verarbeitende Software allerdings, oftmals aus lizenzrechtlichen Gründen, direkt am hostenden Server läuft (problematisch ist dabei die Ablage von Informationen auf einer unverschlüsselten Platte durch die Software selbst, bspw. Bei der Speicherung von Zwischenergebnissen) – selbst im Fall einer komplett zuverlässigen Software muss davon ausgegangen werden, dass Zwischenergebnisse und Datenpartikel auf dem Server zwischengespeichert werden. Speziell komplexe Analysesoftware legt sich beispielsweise oftmals aus Gründen der Performancesteigerung Zwischenergebnisse auf die Platte, weiters spielen Crash-Recovery und Memory-Swapping eine große Rolle.

Abbildung 6 zeigt das Grundkonzept des Ansatzes. Prinzipiell können die Daten hinreichend sinnvoll gelöscht werden, indem der Schlüssel zur VM gelöscht wird. Es ist allerdings zu beachten, dass es jederzeit zu Sicherheitslücken kommen kann, speziell Angriffe auf VM-Systeme haben in den letzten Jahren vermehrt Aufmerksamkeit erhalten. Dabei handelt es sich hauptsächlich um sogenannte Side-Channel-Angriffe, d.h. es wird, oftmals mit Hilfe einer anderen VM am Server, versucht, durch die Analyse und Veränderung von Metainformationen und Metamechanismen (den sog. Seitenkanälen) Informationen über die ausgeführten Befehle und verarbeiteten Daten in einer anderen, am gleichen System gehosteten, virtuellen Maschine zu erlangen. Diese Seitenkanäle basieren oftmals auf Mechanismen des VM-Servers selbst und sind sehr stark vom gewählten Produkt abhängig, bei der Wahl einer derartigen Strategie ist es daher unabdingbar, dass die Entwicklung des hostenden Systems genau beobachtet wird.

*Abbildung 1010 – Virtualisierung*

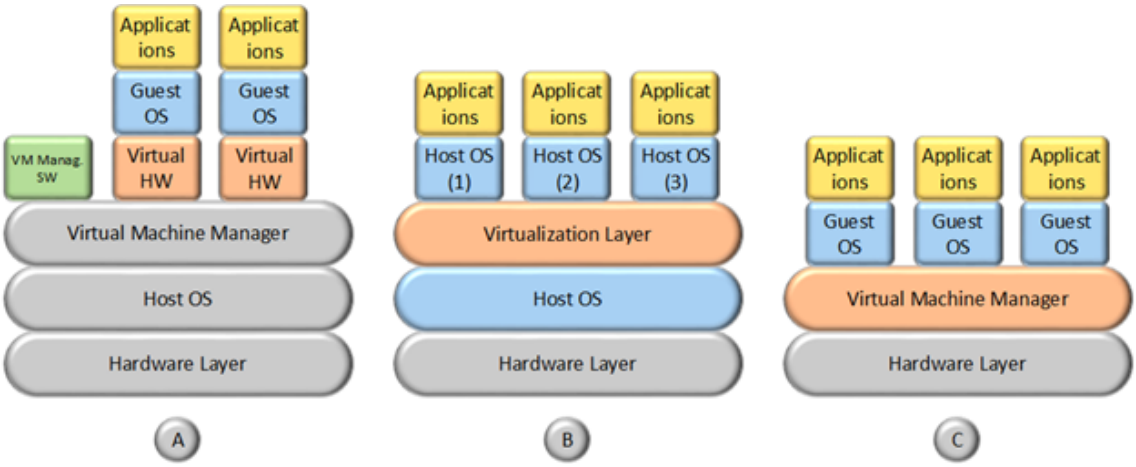



Für die Garantie der Sicherheit ist es essentiell, dass die virtuellen Maschinen vollständig voneinander getrennt sind, d.h. auch Datenverkehr zwischen zwei Maschinen darf nicht direkt durchgeführt werden, sondern immer über eine zentrale, vertrauenswürdige Instanz. Im Folgenden wollen wir kurz einige (klassische) Attacken gegen virtualisierte Umgebungen ansprechen. Typischerweise beinhalten diese Szenarien nicht die Löschung von Daten, sondern vielmehr entweder das Erhalten von (a) irgendeiner Form von Zugriff auf die Daten einer VM, (b) das Ausführen von Code auf einer VM oder (c) Metainformationen (bspw. Ort des Data-Centers). Zusätzlich zu den „normalen" Angriffsszenarien, denen netzwerkbasierte Lösungen grundsätzlich ausgesetzt sind, bieten virtualisierte Lösungen noch ein weiteres Angriffsszenario, gegen das abgesichert werden muss: Ein Angreifer könnte auch ein (legitimer) Nutzer des virtualisierten Systems sein, der (berechtigten) Zugriff auf eine andere virtuelle Instanz besitzt, oder aber sogar, je nach Angriffsszenario und Schutzbedürfnis der Daten, auch der Betreiber des Host-Systems.

Wesentlich für die Möglichkeit der Isolation der einzelnen virtuellen Maschinen ist auch die Art der Virtualisierung. Grob gesprochen kann diese in drei Typen eingeteilt werden (siehe Abbildung 9):

- Full virtualization (A): Auch bekannt als „Virtual Machine Manager" (VMM). Der VMM operiert dabei als Applikation in einem sog. „Host"-Betriebssystem und unterliegt damit auch einer gewissen Kontrolle. Bekannte Produkte sind bspw. VMware Workstation.
- S-layer virtualization (B): Auch bekannt unter dem Namen „Containerbasierte Virtualisierung" basiert darauf, mehrere Instanzen des gleichen Betriebssystems mittels eines sog. VIrtualisierungs-Layers zur Verfügung zu stellen. Bekannte Produkte sind bspw. OpenVZ5, Solaris Container6, Linux VServer8.
- Hardware-layer virtualization (C): Die Idee hinter diesem Konzept ist, dass die einzelnen virtuellen Maschinen direkt auf der Hardware kaufen und lediglich die Ressourcenzuteilung von einem sog. Hypervisor gesteuert wird. Der Hauptvorteil liegt dabei in der starken Isolierung der einzelnen Instanzen und der Effizienz in der Nutzung der Hardware.
-

*Abbildung 1111 – Typen von virtuellen Maschinen*




Neben der Wahl des Virtualisierungskonzepts ist aber das Nutzen von gemeinsamen Ressourcen besonders problematisch:

- Zugriff auf **gemeinsame Datenspeicher**, die nicht extra abgesichert sind. Diese eignen sich sehr gut um bspw. Malware zu verteilen, aber auch zur Exfiltration von Daten aus einer infizierten VM in eine andere (bspw. legal zugreifbare) virtuelle Maschine. Dies ist speziell deswegen sehr relevant, als der Vorgang der Exfiltration, d.h. des Kopierens/Versendens gestohlener Daten ab einer gewissen Datenmenge eine sehr verräterische Angelegenheit sein kann, die einfach durch entsprechende Maßnahmen erkannt werden kann. Zusätzlich ist es auf gemeinsamen Speichern immer wichtig, Zugriffsrechte extrem sorgfältig zu setzen, da sonst der Zugriff auf fremde Daten trivial ist.

- Ein Problem vieler Big-Data-Architekturen, vor allem auch im Cloud-Bereich, lag in der Nutzung von Methoden zur **Daten-Deduplikation**. Die Idee hinter Daten-Deduplikation ist, dass im Fall ähnlicher Anwendungen oftmals die gleichen Daten anfallen. Da Speicher mit gewissen Kosten verbunden ist, ist es natürlich wesentlich kosteneffizienter, statt n-mal den gleichen Datenblock zu speichern, diesen nur einmal auf einem zentralen Speicher zu sichern, zusammen mit einer Liste der n Nutzer, die sich im Besitz dieses Blocks befinden. Techniken wie diese sind auch eine der wesentlichen Argumente für die Nutzung von virtualisierten Systemen, speziell Cloud-environments. Aus dieser, an sich sehr sinnvollen Technik, lassen sich leider, abhängig von der genauen Implementierung und dem Vorwissen des Angreifers, einige Angriffsszenarien generieren:

    - Oftmals gibt es Szenarien, in denen festgestellt werden soll, ob ein gewisser Datensatz vorhanden ist, oder nicht. In einer virtualisierten Umgebung mit Daten-Deduplikation könnte ein Angreifer den zu prüfenden Datensatz selbst in seiner VM erstellen und dann, so er die Bordmittel zur Verfügung hat, feststellen, ob die Daten auch wirklich auf den zentralen Server geladen werden (miss), oder ob nicht (es wird lediglich der Counter des Besitzes des Blocks um eines erhöht – hit). In diesem Szenario ist es eher unwahrscheinlich herauszufinden, welche andere virtuelle Maschine sich noch im Besitz der Daten befindet - das Vorhandensein lässt sich jedoch damit nachweisen. Natürlich ist dieses Verfahren umso schwieriger, je kleiner die Daten sind. Speziell jedoch, wenn es sich um ganze Datensets, oder Datensätze mit sehr großen Argumenten wie bspw. biometrische Daten handelt, sind diese Angriffe jedoch durchaus im Bereich des Möglichen, selbst wenn lediglich rudimentäre Analysewerkzeuge zur Verfügung stehen.

    - Wesentlich ist auch die Frage, wie die Kontrolle über Zugriffsrechte auf Systemen mit Daten-Deduplikation. Manche Systeme, wie bspw. das früher von Dropbox verwendete, basieren auf Hashes: Jede Instanz, die sich bei dem zentralen Speicher meldet, sendet eine Liste an Hashes aller Daten, auf die sie Zugriff hat. Bekommt ein Angreifer Zugriff auf die Liste der Hashes der Daten einer anderen Instanz, so kann er sich dies zu Nutze machen und seien Liste um die erhaltenen Hashes erweitern und wird vom Daten-Deduplikationssystem als legitimer Nutzer dieser Daten gesehen. Im Fall von Systemen, die Daten pushen (wie bspw. Dropbox) würde das System zusätzlich noch aktiv die Exfiltration der Daten übernehmen. Der Vorteil einer solchen Lösung ist allerdings, dass legitime Nutzer ihre Daten jederzeit problemlos replizieren können.

    - Wesentlich ist natürlich auch die Absicherung der Kontroll- und Speicherinstanzen des Deduplikationssystems, speziell in einer zentralisierten Lösung, da diese sehr interessante Angriffsziele sind.

    - Selbst wenn der Zugriff auf die Daten abgesichert ist, bspw. durch eine strikte Kontrolle durch ein zentrales System, das technisch hinreichend abgesichert ist, bietet Daten-Deduplikation einem Angreifer, der es von einer VM auf eine andere geschafft hat (bspw. durch Malware), üblicherweise mannigfaltige Möglichkeiten der Datenexfiltration.

- Ein gerne genutztes Feature in virtualisierten Umgebungen ist das „**Shared Clipboard**“, d.h. die Möglichkeit, Daten von einem VM in die Zwischenablage zu kopieren und diese von einer anderen VM aus wieder auszulesen. Heutzutage wird dies in großen Anwendungen typischerweise nicht mehr genutzt, da bei vielen VMs keine sinnvolle Anwendbarkeit mehr gegeben ist (zu viele Nutzer speichern permanent Daten in die Zwischenablage), manche ältere, oder schlampig konfigurierte, Systeme erlauben dies jedoch.

- **Veraltete Systeme**: Speziell in verschlüsselten virtualisierten Umgebungen kann es zu Problemen kommen, wenn diese nicht entsprechend gewartet werden, aber dennoch Zugriff auf das Internet besteht. Dies passiert vor allem dann, wenn Virtualisierung genutzt wird um Software auszuführen die alte, unsichere und nicht mehr in Wartung befindliche Betriebssysteme nutzt, bspw. Windows XP und dergleichen. Dieses Problem unterscheidet sich nicht grundsätzlich von der Nutzung unsicherer Software mit Netzwerkanschluss im Allgemeinen, allerdings können VMs, die in ressourcenstarken Clustern gehostet werden relativ schnell zu potenten Akteuren in DDOS-Attacken werden. Zusätzlich ist die Nutzung von virtuellen Umgebungen für die Weiternutzung alter Software ein häufig angetroffenes Anwendungsgebiet.

- **Verschlüsselte Kommunikation** zu VMs: Auf der einen Seite bieten verschlüsselte virtuelle Maschinen guten Schutz gegen Ausspähung durch den Netzwerkbetreiber und den Betreiber des Host, bzw. des Hypervisors, auf der anderen Seite haben viele große Betreiber virtualisierter Systeme starke Sicherheitsmaßnahmen, wie Intrusion Detection Systeme (IDS), Firewalls, Virenscanner und dergleichen installiert, die zumindest teilweise durch die Verschlüsselung der Kommunikation hintertrieben werden. Allerdings ist die Nutzung nicht-verschlüsselter Kommunikationsprotokolle die Basis einiger Angriffe, vor allem, wenn der Angreifer bereits Zugriff auf eine andere VM am gleichen Host besitzt.

- **Seitenkanäle**: Oftmals können auch durch sog. Seitenkanäle Informationen zum Status einer virtuellen Maschine in Erfahrung gebracht werden. Bspw. wurde das Zeitverhalten des Sendens von Mails zwischen zwei VMs dazu genutzt um festzustellen, ob die Maschinen im gleichen Data-Center, oder sogar am gleichen Grid lagen. Für sich ist diese Attacke nicht weiter dramatisch, allerdings kann sie zur Vorbereitung weiterer Angriffe genutzt werden. Hierzu gibt es mannigfaltig Literatur in der akademischen Forschung, die bei Bedarf noch extra analysiert werden müsst. Zusätzlich ergeben sich in diesem Gebiet stetig neue Erkenntnisse.

- **Produktspezifische Sicherheitsprobleme**: In den letzten Jahren wurden diverse Angriffe auf unterschiedliche Produkte im Bereich von virtuellen Maschinen und Hypervisoren bekannt, die es teilweise erlaubt, Kommunikation zu tracken, oder gar die Kontrolle über gehostete Maschinen zu erlangen. Virtualisierung erzeugt ein neues Level an Komplexität in der Absicherung und das eingesetzte Personal muss entsprechend geschult werden.

- **Direkter Zugriff** auf **Netzwerke** gepaart mit unsicheren Treibern: Ein Spezialfall der Nutzung alter Systeme, hier greift das Guest-OS direkt auf das Netzwerk zu und

Zusammengefasst kann gesagt werden, dass Virtualisierung gute Lösungskonzepte für die Verarbeitung sensibler Daten bietet, das Aufsetzen einer sicheren virtualisierten Umgebung aber nicht trivial ist. Zusätzlich zu den angesprochenen, sehr generellen, Problemen, kommen noch eine Vielzahl an Konzept- und Produktspezifischen Anforderungen und Parametern, die entsprechend beachtet werden müssen. Dazu kommt noch, dass auch im Fall von Virtualisierung das Sicherheitskonzept auf das Nutzungskonzept abgestimmt werden muss, um sinnvoll in den Analyseprozess integriert werden zu können – oftmals sind es speziell die Anwendungen mit ihren Sicherheitslücken, die eine sichere Lösung hintertreiben.

## 5.3 Anonymisierung von Informationen

Ein wesentliches Problem in der Datenverarbeitung, das speziell in der letzten Zeit an Aufmerksamkeit gewonnen hat, ist das Problem der Sensibilität von Daten und der Schutz der Privatsphäre von Individuen. Dies ist vor allem dann ein wichtiges Thema, wenn es sich bei den sensiblen Daten um personenbezogene Informationen handelt, speziell im Umfeld der medizinischen Datenverarbeitung.

Im Laufe der letzten Jahre wurden dabei verschiedene Methoden des Datenschutzes etabliert, wobei die meisten, wie Verschlüsselung, hauptsächlich gegen Angriffe von außen gerichtet waren, d.h. es ging vorrangig um den Schutz der Informationen vor externen Angreifern. Im Rahmen der Bewusstseinsbildung in Richtung Privacy hat sich jedoch in den letzten Jahren der Trend entwickelt, Daten auch für die interne Analyse als schutzbedürftig anzusehen, d.h. bereits die Verarbeitung möglichst anonymisiert durchzuführen. Dies steht in starkem Einklang mit den aus der DSGV erwachsenden Implikationen. Laut diesen dürfen nicht-anonymisierte Daten nur nach ausdrücklicher Genehmigung durch die Datensubjekte analysiert werden. Dabei ist eine Blankoerklärung nicht ausreichend, es muss der genaue Zweck der Datenanalyse vorher spezifiziert und freigegeben werden (siehe auch die rechtlichen Analysen).

Wesentlich für die Sicherstellung der Anonymität ist dabei eine genaue Analyse der in den Daten enthaltenen Informationen in Hinblick auf die Möglichkeit, aus scheinbar unpersönlichen

Informationen Personen eindeutig identifizieren zu können. Dabei werden die Daten grundsätzlich in drei Typen eingeteilt:

- **Identifier**, also Informationen, die von sich aus eine Person (mehr oder weniger) identifizieren wie Name, SVN und dergleichen. Identifier müssen üblicherweise im Rahmen einer Anonymisierung gänzlich entfernt werden, sie sind aber in den meisten Kontexten eher einfach zu erkennen.

- **Quasi Identifier**, dies sind Daten, die für sich gestellt unproblematisch sind, in Kombination jedoch die Identifizierung ermöglichen. Dies basiert auf einer Studie von Sweeney et. al, bei der festgestellt werden konnte, dass über 70% aller Amerikaner lediglich durch das Tupel (Geburtsdatum, PLZ, Geschlecht) eindeutig identifiziert werden können (Sweeney, (2002b)). Diese Quasi-Identifier sind der Teil des Datensatzes, auf den die meisten Anonymisierungstechniken abzielen.

- **Payload**, oder Nutzdaten, dies sind diejenigen Daten, die in Hinblick auf persönliche Identifizierbarkeit unproblematisch sind und oftmals den für die Analyse interessantesten Teil ausmachen, wie bspw. Diagnosedetails. Speziell bei der Verschränkung von Nutzungsdaten mit sozioökonomischen Informationen muss jedoch speziell darauf geachtet werden, letztere nicht voreilig als unproblematische Payload zu definieren, da diese Daten oftmals in Kombination mit anderen Merkmalen weitreichende Rückschlüsse, oder gar die De-anonymisierung von Individuen erlauben.

Abbildung 8 gibt einen Beispieldatensatz an, in dem auch die unterschiedlichen Informationstypen beispielhaft dargestellt sind.

Je nach Angreiferpotenzial und konkreter Gestalt der Daten wurden dabei unterschiedliche Anonymisierungsverfahren entwickelt. Auf den folgenden Seiten besprechen wir kurz die derzeit gebräuchlichsten Algorithmen und deren inhärenten Problematiken:

- Die Anonymisierung mit Hilfe von synthetischen Daten, Katastern und dergleichen.

- K-anonymity und davon abgeleitete Verfahren, sowie mehrere Methoden zur Erstellung k-anonymisierter Datensätze. Dieses Verfahren ist speziell in der akademischen Forschung von großer Bedeutung.

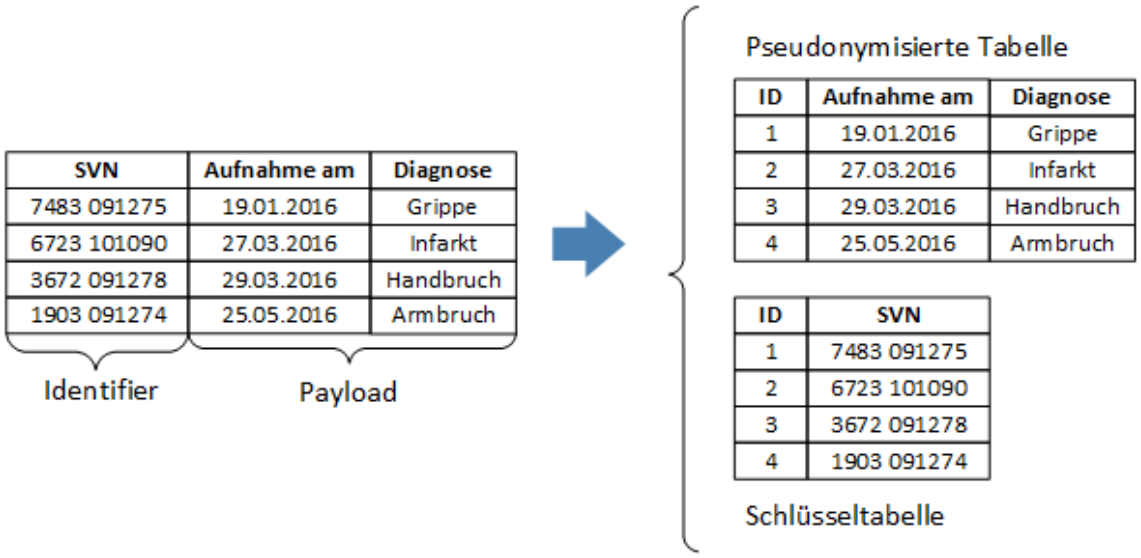
- Differential Privacy, als relativ neuer orakelbasierter Ansatz, der ebenfalls großen Widerhall in der Forschung erfährt.

Eines der Hauptprobleme beim praktischen Einsatz von Anonymisierungsverfahren ist das Fehlen – auch bei der DSGV - exakt definierter rechtlicher Anforderungen an die Stärke der Anonymisierung (entspricht im Fall von k-anonymity dem Faktor k der Mindestgröße der Äquivalenzklassen). Zusätzlich ist eines der wesentlichen Probleme der meisten Anonymisierungsverfahren die Inferenz mit anderen Daten: Die Verfahren gehen davon aus, dass der Angreifer lediglich Zugriff auf die anonymisiert zur Verfügung gestellten Daten hat. Problematisch wird diese Annahme, wenn er auch Zugriff auf verwandte Datensätze besitzt, oder externes Wissen über die darin enthaltenen Personen mitbringt. Speziell in Hinblick auf letzteres Problem kann es keine allgemeine Lösung technischer Natur geben, da sich immer ein trivialer Datensatz als Extrembeispiel generieren lässt, der eine Anonymisierung aufhebt. Wie auch im Bereich der Kryptographie wird daher auch im Bereich der Anonymisierung eine etwas reduzierte Anforderung der Nutzung von Algorithmen und Anonymisierungsstärken als Basis des derzeitigen State-of-the-Arts angenommen werden müssen.

### 5.3.1 Pseudonymisierung

Grundsätzlich besitzen Methoden der Pseudonymisierung andere Ziele als Anonymisierungsverfahren. Wie letztere zielen sie darauf ab, dass die Verarbeitung von Daten ohne die Preisgabe identifizierender Informationen durchgeführt werden kann, wodurch sie oft zu den Anonymisierungsverfahren gezählt werden. Der wesentliche Unterschied liegt darin begründet, dass Pseudonymisierung darauf abzielt, die Daten nach der erfolgten Verarbeitung wieder zu De-pseudonymisieren. Dazu werden die Daten aufgeteilt, die identifizierenden Daten werden aus der Tabelle entfernt und in eine eigene Tabelle gespeichert, die Ursprungsdaten werden quasi in zwei Tabellen gespaltet, von denen die eine für die weitere Verarbeitung genutzt wird, während die andere die identifizierenden Daten enthält. Ein Schlüssel („ID“ im Beispiel in Abbildung 10) sorgt dafür, dass die Datensätze wieder zusammengeführt werden können. Ein weiteres wesentliches Problem liegt in den Quasi-Identifiern begründet. Da diese geeignet sind, eine Person auch ohne die Schlüsseltabelle zu identifizieren, müssen diese ebenfalls aus der pseudonymisierte Tabelle in die Schlüsseltabelle entfernt werden. Jedoch sind es oftmals gerade diese Informationen, die im Rahmen einer statistischen Auswertung von Interesse wären, bspw. Ist die Frage nach Alter oder Geschlecht in medizinischen Analysen oftmals hochrelevant.

*Abbildung 12 – Pseudonymisierungsbeispiel*




Die Erstellung des Schlüssels (ID) kann dabei auf verschiedene Arten gewählt werden. Typisch, aber nicht erschöpfend, werden oftmals die folgenden Techniken angewandt:

- Systematische Schlüssel, bspw. aufsteigend nach Reihenfolge der Daten in der Tabelle
- Schlüssel, die abhängig von Werten aus der Schlüsseltabelle. Hierbei werden oftmals Hashwerte eingesetzt.
- Geheime Schlüssel, als Beispiel kann hier die Bereichsspezifische Verarbeitungskennzahl dienen.
- Zufällig gewählte Schlüssel.
-

Rechtlich gesehen werden pseudonymisierte Daten wie sensible, unanonymisierte Daten behandelt (siehe auch die rechtlichen Ausarbeitungen zu dem Thema der Pseudonymisierung). Aus diesem Grund werden wir uns in weiterer Folge nicht weiter mit den verschiedenen Methoden der Pseudonymisierung befassen.

Dennoch können Pseudonymisierungsverfahren einen wertvollen Schutz der Daten leisten, selbst wenn sie dadurch nicht besser gestellt im Sinne der GDPR werden: Pseudonymisierung als Sicherheitsmaßnahme erfreut sich speziell in der medizinischen Forschung großer Beliebtheit, da ohne den Zugriff auf die Schlüsseltabelle typischerweise kein Personenbezug hergestellt werden kann (dies trifft natürlich nur eingeschränkt zu, sollten die in der Auswertung vorliegenden Quasi-Identifier ausreichen um die Person auch auf Basis dieser Daten identifizieren zu können). Dadurch hat man einen sehr effizienten Sicherheitsmechanismus an der Hand, der keinerlei weiter Verzerrung in der Datenauswertung einführt und entsprechend das Sicherheitsniveau, und damit die Sorgfalt im Umgang mit sensiblen Daten, hebt. Liegt Consent für die spezifische Auswertung vor, so ist die Anwendung von Pseudonymisierungsverfahren ein wesentliches Instrument zum Schutz des Personenbezugs und ein wichtiges Mittel zur Erhöhung der Informationssicherheit.

Wesentlicher Aspekt aller Datenanwendungen unter dem Regime der GDPR ist aber immer die Frage, ob die gesammelten Daten überhaupt wesentlich für die geplante Analyse sind. Hierbei kann in vielen Auswertungen die Sensibilität der Daten deutlich reduziert und oftmals der Personenzeug vermieden werden, indem die in der GDPR geforderte Datensparsamkeit beachtet wird und sensible Informationen, die oftmals für, speziell statistische, Auswertungen nicht notwendig sind, entsprechend vor der Auswertung, oder besser noch zum Zeitpunkt des Sammelns, entfernt werden.

### 5.3.2 Datenperturbation, Aggregation und Kataster

Die grundsätzlich einfachste Methode der Anonymisierung großer Datensätze stellt die Verfälschung der Daten durch Methoden der Perturbation, oder aber der Zusammenfassung in Datenklassen dar. Dabei gibt es mehrere Möglichkeiten, von denen hier einige angesprochen werden:

- ***Perturbationsmethoden*** wie (stochastische) Überlagerung numerischer Informationen, Vertauschung von Merkmalen zwischen den Datensätzen, sowie die Anreicherung mit zusätzlichen Daten, die entweder synthetisch kreiert wurden, oder unbeteiligte Personen aus der Grundgesamtheit darstellen. Hierbei gibt es verschiedene Angriffspunkte, die in der Literatur entsprechend diskutiert werden. Typischerweise hängt die genaue Wahl einer Perturbationsmethode von weiteren Rahmenbedingungen ab, bspw. den Erhalt statistischer Eigenschaften.

- ***Aggregation*** kann ebenfalls genutzt werden, um die Identifizierung Einzelner aus einer Datenmenge zu erschweren. Dabei werden überhaupt keine Einzeldatensätze mehr angegeben, sondern nur mehr (u.U. statistische) Auswertungs- und Aggregierungsergebnisse, wie bspw. durchschnittliche Verdienste oder Krankheitsfälle aufgeschlüsselt auf Regionen, die es nicht mehr erlauben auf den Einzeldatensatz rückzuschließen. Grundsätzlich kann natürlich immer diskutiert werden, ob nicht bspw. allein die Herkunft einer Person bspw. aus einem Gebiet mit stark erhöhtem Krebsrisiko bereits ebenfalls eine sensible Information darstellt.

- ***Kataster***: Zusätzlich muss bei Aggregierungsverfahren das Problem wiederholter Abfragen mit unterschiedlichen Aggregierungskriterien im Auge behalten werden. Bspw. kann durch die Durchführung der gleichen Auswertung (Durchschnittsgehalt pro geographischem Gebiet) durch die leichte Veränderung der Aggregationsgebiete das Gehalt einer bestimmten Entität berechnet werden. Kataster dienen dazu, die

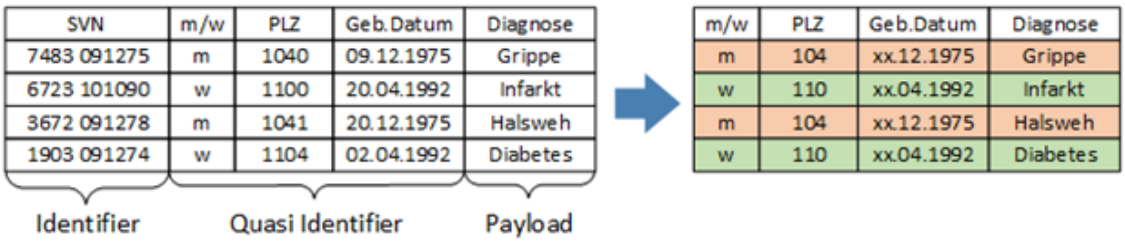
Aggregationsgebiete a-priori zu fixieren, d.h. dieser Angriff ist nicht mehr möglich, da die Grenzen der aggregierten Gebiete immer gleich bleiben. Am besten kann man sich das mit geographischen Gebieten vorstellen, die Aggregationsebenen wie bspw. Block, Gemeinde, Großgemeinde, Bezirk bleiben immer gleich.

Alle drei vorgestellten Techniken sind allerdings problematisch in Hinblick auf dynamische Daten, d.h. wenn die Anonymisierung auf sich verändernde Datenmengen angewandt werde muss. Dies ist allerdings derzeit noch ein generelles Problem im Bereich der Anonymisierung, das auch auf andere Verfahren zutrifft (Xiao et. al., (2007)).

### 5.3.3 K-anonymity und abgeleitete Verfahren

Das Verfahren der k-anonymity (Sweeney, (2002)) beruht darauf, die Quasi-Identifier so zu verändern, dass mindestens k Datensätze in der gleichen Äquivalenzklasse liegen, d.h. in Bezug auf die Quasi Identifier gleich sind (siehe Abbildung 11). Der Faktor k wird dabei auch Anonymisierungsfaktor genannt und mit der Stärke der Anonymisierung identifiziert.

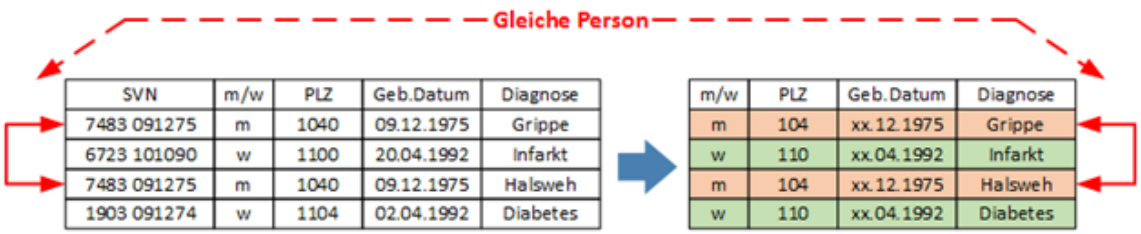
*Abbildung 1313: k-anonymisierter Datensatz mit k=2*




Klassische k-anonymity besitzt allerdings einige Schwächen, die wir im Folgenden diskutieren werden. Im Laufe der Jahre wurde das Konzept der k-anonymity daher immer weiter erweitert, bis eine Familie verwandter Anonymisierungskonzepte entstand, die teilweise sehr spezifisch auf eng umrissene Probleme und deren Rahmenbedingungen angepasst sind. Wir werden uns daher im Folgenden lediglich auf die wichtigsten konzentrieren.

Ein Standardproblem im medizinischen Bereich ist die Tatsache, dass einer Person durchaus mehrere Datensätze in einer medizinischen Datenbank zugeordnet werden können. Dies ist speziell bei stationären Aufenthalten eher Standard als Ausnahme. Das Problem dabei ist, dass das Konzept der k-anonymity auf der impliziten Annahme beruht, dass jeder Datensatz genau einer Person zugeordnet ist. Ist dies nicht der Fall, so werden alle den gleichen Patienten betreffende Datensätze (so sich seine Quasi Identifier nicht geändert hatten) automatisch der gleichen Äquivalenzklasse zugeordnet, dadurch wird selbst ein hoher Faktor k relativ schnell erfüllt, obwohl nicht Datensätze k unterschiedlicher Personen pro Äquivalenzklasse enthalten sind. Abbildung 12 zeigt ein entsprechendes, stark vereinfachtes Beispiel.

*Abbildung 1414 – k-anonymity mit multiplen Datensätzen*



Zur Umgehung des dargestellten Problems wird daher in Datenmengen, die multiple Datensätze enthalten können, die Zusatzforderung eingebracht, dass nicht k Datensätze pro Äquivalenzklasse, sondern die Datensätze von k unterschiedlichen Personen pro Äquivalenzklasse enthalten sein müssen, die sog. (X,Y)-anonymity (Yang et. al., (2006)).

Ein weiterer Nachteil an dem Konzept der k-anonymity liegt darin begründet, dass eine Deanonymisierung eines spezifischen Datensatzes unter Umständen gar nicht benötigt wird um heikle Daten zu extrahieren. Angenommen sei bspw. eine Datenbank an Erkrankungen. K-Anonymity stellt sicher, dass es nicht möglich ist, einen Datensatz in der Tabelle genau einer einzigen Person zuzuordnen, d.h. die Personen sind ununterscheidbar den Datensätzen zugeordnet. Kann der Angreifer aber eine Person zumindest einer Äquivalenzklasse zuordnen, so kann er dadurch schon ein gewisses Zusatzwissen generieren, speziell wenn die Erkrankungen innerhalb der Äquivalenzklasse gleich sind, dann wurde sogar der gleiche Effekt wie bei einer De-Anonymisierung erreicht, bspw. ist es unerheblich wirklich die Datensätze eindeutig einem Nutzer zuordnen zu können, wenn die Erkrankung für alle in Frage kommenden Datensätze gleich ist. Das Konzept der l-diversity sieht daher vor, dass die Äquivalenzklassen in den einzelnen Merkmalen immer mindestens einen verschiedenen Wert aufweisen (Machanavajjhala et. al., (2007)).

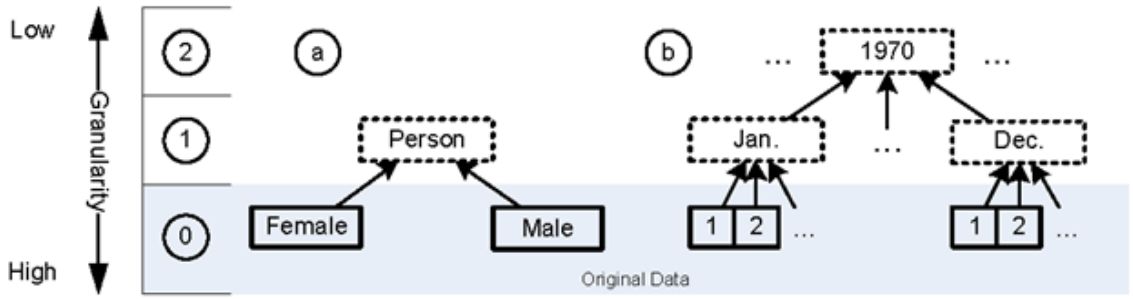
Die Erweiterung des Konzepts zur t-closeness verlangt sogar, dass die Verteilung der Merkmale mit der der Originaltabelle übereinstimmt (Li et. al., (2007)). Wesentlicher Nachteil dieses Konzepts ist, dass es oft unmöglich ist noch einen sinnvoll nutzbaren Datensatz zu erstellen, der t-closeness mit einem gewissen Schutzfaktor erfüllt. Praktische Anwendungen sind uns derzeit nicht bekannt.

### 5.3.4 Methoden zur Generierung von k-anonymen Datensätzen

In der Literatur wird k-Anonymity oftmals als Generalisierungsmethode bezeichnet, d.h. die Granularität der Information in den Quasi-Identifiern wird solange reduziert, bis das k-Anonymity-Kriterium erfüllt ist. Dies ist allerdings nicht ganz korrekt, da Generalisierung nur eine (allerdings die dominante) Methode zur Erzeugung k-anonymer Datensätze darstellt. Zusätzlich wird mit Generalisierung typischerweise „Full-domain Generalisierung“ gemeint, auch hierzu gibt es einige Erweiterungen. Die Ergänzung von Generalisierung durch andere Methoden, sowie die Wahl komplexerer Generalisierungsmethoden hat den Hintergrund, dass Full-domain-Generalisierung sehr stark von der Fragmentierung der Daten im Wertebereich abhängt, d.h. einzelne Ausreißer in einem Kriterium können die Qualität der generalisierten Daten wesentlich beeinflussen (wobei auch die Messung der Datenqualität mit sog. data precision metrics ein nichttriviales Thema darstellt).

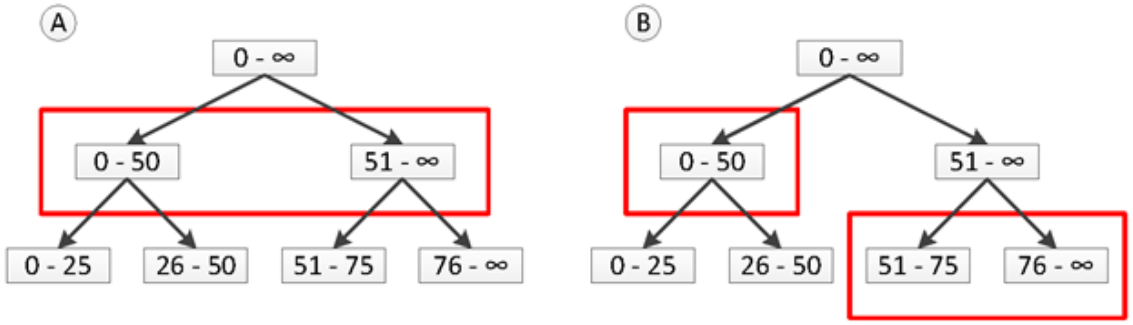
Das Prinzip der Full-Domain-Generalisierung basiert dafür, die Granularität der Quasi-Identifier (einer Spalte, eines Attributs, oder welche Bezeichnung auch immer gewählt wird) solange zu reduzieren, bis das k-anonymity-Kriterium erfüllt ist. Dies kann natürlich auf verschiedene Arten geschehen, typischerweise wird eine sogenannte Data Precision Metric zur Messung des Datenqualitätsverlusts verwendet. Diese Metriken können sehr allgemein und simpel sein, oder aber auch spezifisch auf die Wichtigkeit der einzelnen Attribute eingehen (spezielle Metriken im medizinischen Bereich). Abbildung 13 zeigt beispielhaft Generalisierungsstrategien für zwei verschiedenen Quasi-Identifier (QI) „Geschlecht“ und „Geburtsdatum“. Bei der Full-Domain-Generalisierung ist zu beachten, dass ein QI in allen Datensätzen gleichbehandelt wird, d.h. wird das Geburtsdatum auf Monat genau granuliert (Ebene 1 in Abbildung 10), so wird dies für jeden einzelnen Datensatz genau gleichgemacht.

*Abbildung 1515 – Generalisierungsstrategien für 2 QIs*




Letzterer Aspekt kann natürlich zu einem sehr großen Qualitätsverlust führen, vor allem, wenn die Daten sehr ungleich verteilt sind. Eine Gegenstrategie dagegen besteht darin, dass nicht alle Äste im Generalisierungsbaum mit der gleichen Granularität generalisiert werden, sondern Unterbäume unterschiedlicher Granularität definiert werden, die sog. Subtree-Generalization. Abbildung 11 zeigt ein einfaches Beispiel zur Generalisierung des QIs „Alter“ (mit einer zugegebenermaßen sehr optimistischen oberen Schranke). Teil (A) zeigt normale Full-Domain-Generalisierung, (B) eine mögliche Strategie der Subtree-Generalisierung. Dabei ist zu beachten, dass es sich hierbei um den gleichen QI handelt, der je nach Wertebereich unterschiedlich generalisiert wird. Die Berechnung solcher Subtree-Generalisierungen ist natürlich deutlich aufwändiger, da mehr valide Fälle durchgerechnet werden müssen, sie bietet aber immer mindestens genauso gute Ergebnisse wie Full-Domain (die als Spezialfall gesehen werden kann) und typischerweise bessere Ergebnisse auf nichtgleichverteilten QIs.

*Abbildung 1616 – Subtree Generalization*




Auch dieser Ansatz kann noch etwas optimiert werden, zur sog. „Sibling-Generalization“ (siehe Abbildung 15, Seite (B)). Dabei werden für einen Subtree noch spezifische Ausnahmen definiert, d.h. der Wertbereich von 0 bis 50 wird grundsätzlich in das Intervall [0;50] generalisiert, außer die Person ist unter 19 Jahre alt, dann wird das Subintervall [0;18] ausgewählt. Auch hierbei gilt, dass der Aufwand der Erstellung größer ist, das Ergebnis aber mindestens so gut wie im Fall der Subtree-Generalization ist.

*Abbildung 1717 – Sibling Generalization*

Generalisierungsstrategien können allerdings immer entarten, d.h. sehr schlechte Ergebnisse produzieren, sollten einzelne sog. Outlier, d.h. Datensätze die komplett anders als der Rest in einem QI sind, vorhanden sein. In diesem Fall kann es zielführend sein, Datensätze zu löschen, bevor die Generalisierung durchgeführt wird. Dieses Verfahren wird „Suppression“ genannt und besitzt den Vorteil, dass Outlier einfach in einem vorgelagerten Schritt behandelt und entfernt werden können, allerdings gibt es auch einen Nachteil, der je nach Anwendungsfall unterschiedlich zu bewerten ist: Die entfernten Datensätze haben keinerlei Einfluss mehr auf die Analyse, d.h. speziell wenn nach seltenen Phänomenen gesucht wird, kann die Entfernung bereits einer sehr geringen Menge an Datensätzen zu extremen Verfälschungen führen. Dies ist speziell dann der Fall, wenn das Phänomen in einem Zusammenhang mit dem Outlier-Dasein der entfernten Datensätze liegt. Daher ist die Verwendung von Suppression immer mit sehr großer Vorsicht zu genießen, um nicht durch das Artefakt der Entfernung falsche Aussagen aus den Daten herauszulesen.

### 5.3.5 Differential Privacy

Oftmals will man einem Partner nicht den ganzen Datensatz anonymisiert weitergeben, sondern man möchte allgemeine, statistische Informationen zugänglich machen. Differential Privacy (Dwork, (2008)) verfolgt dabei einen stark orakelbasierten Ansatz, was bedeutet, dass nicht die Gesamtdatenmenge in anonymisierter Form an die Analysten weitergereicht wird, sondern die Analysten Abfrage (bspw. Statistische Auswertungen) an die Datenbasis stellen und die entsprechenden Ergebnisse so verzerrt werden, dass keine eindeutigen Rückschlüsse auf Personen möglich sind. Dies ist speziell in Hinblick auf die rechtlichen Rahmenbedingungen problematisch, da die Daten selbst nicht anonymisiert, sondern im Gegenteil in maximaler Genauigkeit vorgehalten werden müssen.

Es ist daher noch zu klären, in welchen Szenarien Differential Privacy überhaupt rechtlich sinnvoll einsetzbar ist.

### 5.3.6 Löschen und Anonymisierung bei Machine-Learning

Eine wesentliche Frage bei der Beurteilung des Themas Datenschutz auf die Innovationskraft von Unternehmen, sowie die Möglichkeit der Entwicklung neuer, datengetriebener Services und Produkte, ist der Effekt der getroffenen Maßnahmen auf die der Datenauswertung zugrundeliegenden Algorithmen. Dabei muss beachtet werden, dass speziell in selbstlernenden, sog. „intelligenten“ Systemen, die Daten nicht nur verarbeitet werden, sondern auch in einem wesentlichen Ausmaß das System konstituieren, d.h. bereits verarbeitete und klassifizierte Daten dienen als Grundlage zur weiteren Analyse, als sog. „Wissensbasis“.

Die Verfälschungseffekte, die durch eine Modifikation an der Wissensbasis auftreten, können daher wesentlich für die weitere Verarbeitung sein, allerdings stand dieses Thema bisher kaum im Fokus der wissenschaftlichen Forschung. Speziell die unterschiedlichen Effekte einer Löschung von Datensätzen wurde bisher nach unserem Dafürhalten, erstmals im Rahmen einer Publikation (Malle et. al., (2016)) im Rahmen der Durchführung des KIRAs-Projekts „DIANGO“ (Digitale Informationsvisualisierung aus automatisierter Analyse von Nachrichten,

Geoinformation und multimedialen Objekten)[10] ausführlich untersucht und miteinander verglichen. Dazu wurden zwei grundlegende Experimente durchgeführt:

- Simulation des Rechts auf Vergessenwerden auf einer Datenmenge, wobei zusätzlich noch ein Worst-Case in Hinblick auf die Wichtigkeit der Merkmale angenommen wurde.
- Die Anonymisierung der Daten vor der weiteren Verarbeitung mit Hilfe von k-anonymity, wobei unterschiedliche Stärken für k angenommen werden.

Auf den so erstellten Datensätzen wurden klassische Machine-Learning-Algorithmen ausgeführt und untersucht, welche Verzerrung durch die jeweiligen Maßnahmen in der Auswertung entstanden. Als Maß der Verzerrung wurde dabei der F1-Score (F-Maß) verwendet, der die Genauigkeit (precision) und die Trefferquote (recall) eines Algorithmus auf einem Datensatz mit Hilfe des gewichteten harmonischen Mittels kombiniert:

Der F1-Score ist ein extrem weit verbreitetes Maß für die Qualität von Ergebnissen, da er auch für sehr ungleich verteilte Daten realistische Ergebnisse liefert. Dabei wurden, unter anderem, die folgenden Algorithmen evaluiert, die zu den wichtigsten und am weitesten verbreiteten Algorithmen im Bereich des Machine Learnings zählen:

- Gradient Boosting
- Linear SVC
- Logistic Regression
- Random Forest

Das zu untersuchende Datensample bestand dabei aus amerikanischen Census-Daten, die in der ML-Forschung sehr häufig als Referenzdatenmenge herangezogen werden. Dabei handelt es sich um Daten die der Forschung zur Verfügung stehen (open data Initiative). Um einen möglichst großen Effekt des Rechts auf Vergessenwerden zu simulieren (Worst Case) wurde eruiert, welche Datenspalte (also welche Merkmale) für den jeweiligen gewählten Machine-Learning-Algorithmus für die Evaluierungsfragestellung am wertvollsten sind. Eine der drei wichtigsten Merkmale wurden dann jeweils aus dem Datensamples entfernt, beginnend mit 20% bis hin zu 100% in 20%-Schritten, d.h. im ersten Lauf wurden aus der Datenmenge für den zu untersuchenden Algorithmus bei 20 Prozent der Datensätze die Daten aus einem der drei wichtigsten Merkmalen entfernt, anschließend aus 40%, 60% und 80%, bis hin zu einer kompletten Löschung des Merkmals aus allen Datensätzen. Dies wurde für die drei wichtigsten Merkmale durchgeführt, was zu 15 neuen Datensätzen führte. Dabei ist zu beachten, dass die Auswahl der Datensätze, aus denen die Merkmale entfernt wurden, zufällig gewählt wurden, d.h. hierbei kann natürlich diskutiert werden, ob Personen, die das Recht auf Vergessenwerden zur Anwendung bringen, spezielle Eigenheiten besitzen, die mehr Einfluss auf die weitere Verarbeitung besitzen.

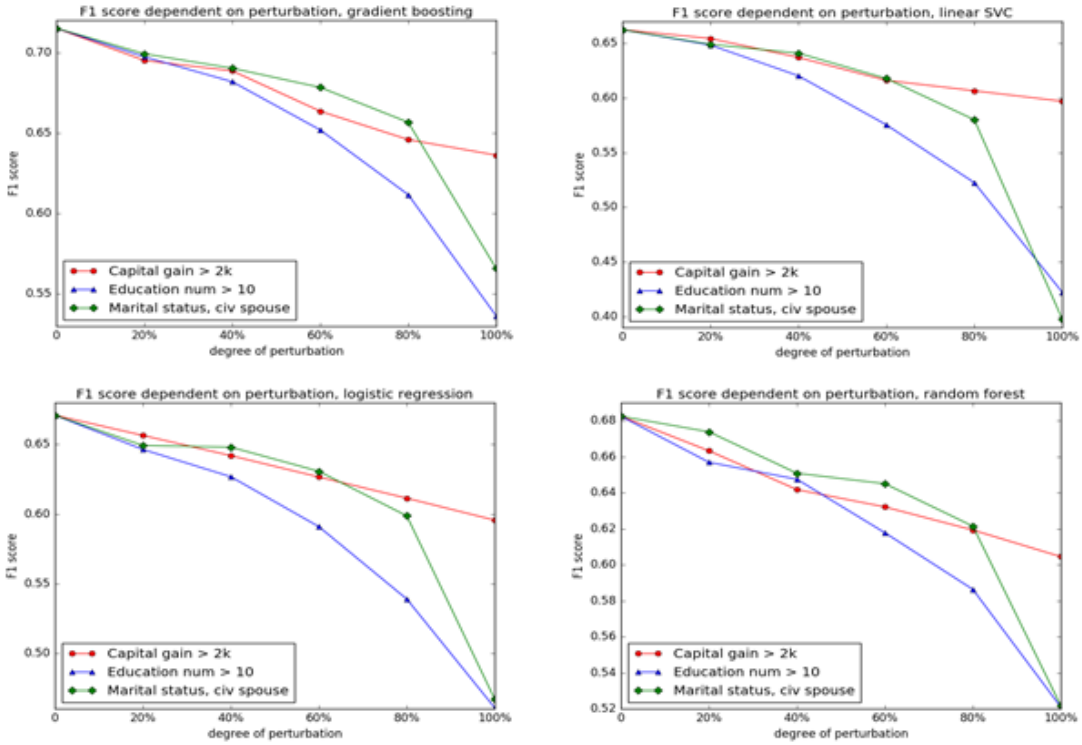
Abbildung 16 gibt einen guten Überblick über den Effekt dieser Entfernung, d.h. der Exekution des Rechts auf Vergessenwerden. Dabei ist zu beachten, dass selbst die kleinste Versuchsanordnung bereits bei 20% liegt, allerdings, und dies ist Gegenstand weiterer Forschungen, wurden nicht die gesamten Datensätze aus dem Sample entfernt, sondern nur die

[10] http://www.kiras.at/gefoerderte-projekte/detail/?tx_ttnews%5Btt_news%5D=325&cHash=bf752b02238ba0209a343753cdb2a5f6

wichtigsten Merkmale. In derzeit stattfindenden Versuchsreihen wird dieses Experiment noch entsprechend verfeinert und realistischer gestaltet. Wir rechnen mit entsprechenden Ergebnissen im Laufe des nächsten Quartals.

Um eine bessere Vergleichbarkeit zu erreichen wurde auch jeweils der F1-Score des originalen Datensatzes mit einbezogen. Es kann gut erkannt werden, dass die Ergebnisse zwar schlechter werden, dies aber überraschenderweise relativ langsam geschieht, d.h. dass selbst bei rund 40% veränderten Datensätzen noch relativ gute Ergebnisse erzielt werden können.

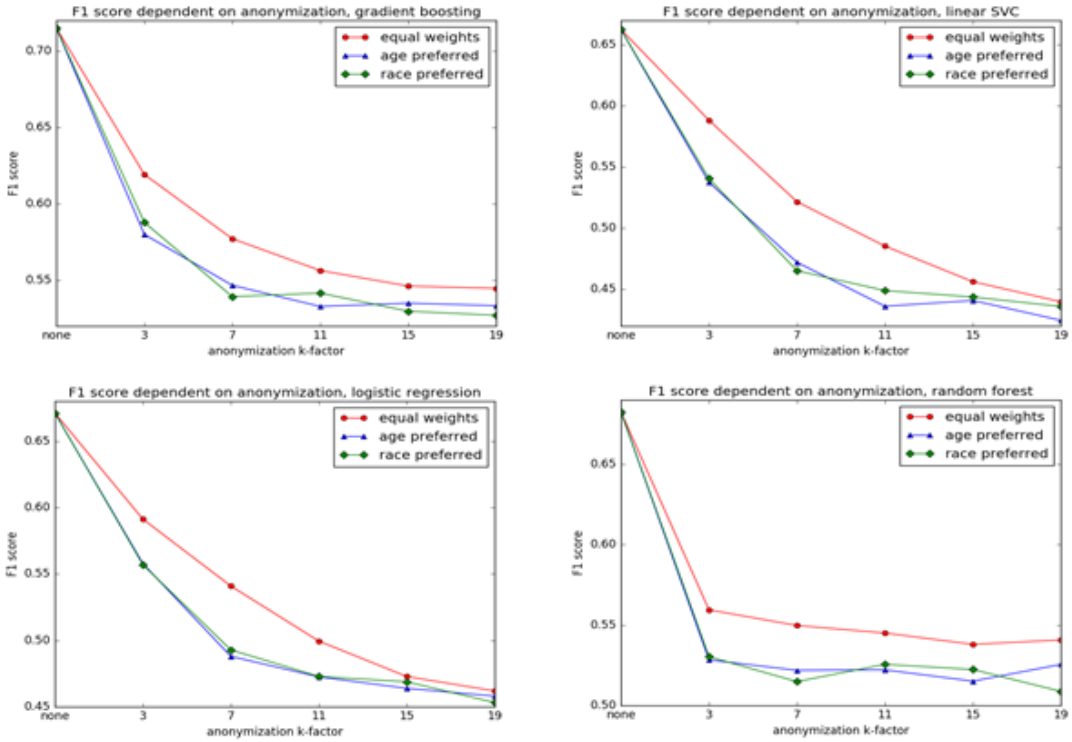
*Abbildung 1818: Selektives Löschen wichtiger Merkmale*




Um den Effekt der Anonymisierung zu studieren, wurde der gleiche Ursprungsdatensatz als Ausgangspunkt gewählt. Die Daten wurden dann k-anonymisiert, wobei die Werte untersucht wurden. Zusätzlich wurde das Clustering noch gewichtet um im Fall der Zuordenbarkeit eines Datensatzes zu verschiedenen Clustern (siehe Abschnitt zu k-anonymity) besser steuern zu können

Abbildung 17 zeigt die Ergebnisse. Im Gegensatz zur Löschung sensitiver Merkmale aus den Datensätzen hat die Anonymisierung sehr großen Effekt auf den F1-Score in unserer Untersuchung, Anonymisierung hat in unserem Beispiel also einen weitaus größeren Effekt auf die Nutzbarkeit der Daten, als die selektive Löschung wichtiger Merkmale. Allerdings muss in Hinblick auf die Einschränkungen und Rahmenbedingungen der Ergebnisse zur Löschung darauf hingewiesen werden, dass auch in diesem Bereich noch weitaus mehr Forschung betrieben werden muss, speziell in Hinblick auf die zu wählende Anonymisierungsmethode. Dennoch bietet dieses Ergebnis einen ersten Anhaltspunkt, in welche Richtung noch speziell Forschungsaufwand benötigt wird, um Techniken des Privacy Aware Machine Learnings (PAML) zu entwickeln, die mit herkömmlichen Techniken mithalten, oder diesen qualitativ zumindest nahekommen können.

*Abbildung 1919: Effekt der k-Anonymity*




## 5.3.7 Zusammenfassung

Grundsätzlich kann gesagt werden, dass auf Basis der GDPR ein Vorgehen, das auf die Sammlung personenbezogener Daten verzichten kann, bspw. durch den Einsatz von Statistiken und der reinen Sammlung von Aggregaten, auf jeden Fall wesentliche Probleme zu vermeiden hilft. Dies trifft auch direkt die explizite Forderung der GDPR nach Datensparsamkeit. Davon abgesehen, kann im Fall des Vorliegens von Consent Pseudonymisierung wertvolle Dienste leisten, um die Datensicherheit wesentlich zu erhöhen, da mit Hilfe dieser Methoden direkte Identifizierbarkeit in indirekte Identifizierbarkeit zu verwandeln, bzw., wenn keinerlei ausreichende QIs in den pseudonymisierten Daten vorliegen, die Re-Identifizierbarkeit wesentlich vom Wissen um die Schlüsseltabelle abhängig zu machen.

Werden personenbezogene Daten in der weiteren Auswertung benötigt und liegt keinerlei Consent vor so sind die Daten zu anonymisieren, wobei natürlich, wie gezeigt, teilweise erhebliche Verzerrungen bei der Nutzung von Statistiken oder Machine-Learning entstehen können

# 5.4 Privacy und Transparenz

Die Forderung nach einer transparenten Verarbeitung der Daten entspringt direkt der DSGV und ermöglicht damit dem Besitzer der Daten eine sehr weitreichende Kontrolle über die Verwendung der Informationen. Zusätzlich ermöglicht sie, technisch gesehen, auch die Kontrolle, ob wirklich nur die angegebenen Daten und Informationen für eine datengetriebene Anwendung verwendet wurden. Allerdings kann diese Forderung auch mit dem Recht auf Vergessenwerden kollidieren, speziell wenn die Forderung nach Transparenz aufgrund anderweitigen Regularien begründet wird. Regularien wie SOX (Sarabanes Oxley Act) und Basel II stellen die Integrität der in einer

datengetriebenen Verarbeitung verwendeten Daten sicher, d.h. sie stellen sicher, dass die verwendeten Daten zu jeder Zeit garantiert nicht manipuliert wurden. Dies gilt speziell auch für externe Anreicherungsinformationen, sodass hiermit auch ein Reprocessing (Nachverarbeitung) ermöglicht wird, d.h. es ist möglich Daten so zu verarbeiten, wie das an einem bestimmten Zeitpunkt mit den damals vorliegenden Informationen gemacht worden wäre. Dies ist speziell wichtig in Billing-Workflows und generiert eine gewisse Beweisbarkeit gegenüber Forderungen und Anfechtungen, ist daher speziell im Bereich der Finanztransaktionen von hoher Bedeutung.

Eine hohe Transparenz ermöglicht typischerweise allerdings keine Entfernung von (verarbeitungsrelevanten) Informationen, ohne zumindest die Nachverarbeitbarkeit einzuschränken. Daher sollte im Rahmen der Forderung der Transparenz die folgende Unterscheidung getroffen werden:

- Transparenz in Hinblick auf die Garantie der korrekten Verarbeitung zum Verarbeitungszeitpunkt
- Transparenz im Sinne der Garantie der Sicherheit gegenüber nachträglicher Manipulation
- Transparenz, die eine Nachverarbeitung ermöglicht.

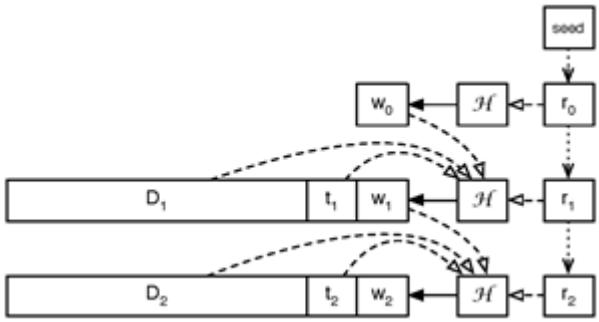
Die erste Forderung ist eine relativ schwache Forderung, bei der eine sog. „trusted entity“ (vertrauenswürdige Entität) garantiert, dass die verwendeten Daten zu diesem Zeitpunkt korrekt waren, ohne aber Details zu diesen Daten preiszugeben. Diese Forderung kann relativ einfach mit Hilfe von digitalen Signaturen umgesetzt werden, die Entity signiert die Ergebnisse und den entsprechenden Zeitstempel mit ihrem „Private Key“, speichert diese am System und bürgt damit für die korrekte Durchführung zum Verarbeitungszeitpunkt. Eine kontrollierte Nachverarbeitung ist in diesem Fall ausgeschlossen, auch muss sehr viel Vertrauen in diese Entität gelegt werden. Zusätzlich muss diese überhaupt in die Lage gebracht werden, die Richtigkeit der Daten kontrollieren und garantieren zu können. Ein weiteres Problem bei dieser relativ naiven Technik liegt darin begründet, dass signierte Ergebnisse von einem Angreifer mit entsprechenden Privilegien am System relativ einfach entfernt werden können, da die einzelnen Ergebnisse quasi unabhängig und unverkettet vorliegen. Auch können nachträglich neue Ergebnisse hinzugefügt werden. Ein Angreifer, der sich im Besitz des privaten Schlüssels der trusted entity befindet (oder auch eine korrumpierte trusted entity selbst), kann nicht genehme Ergebnisse entfernen und diese durch nachträglich erstellte Ergebnisse mit alten Zeitstempeln ersetzen. Durch dieses Problem, aber auch wegen der Nichtnachvollziehbarkeit der Anreicherungsschritte selbst (es wird ja nur das Ergebnis quasi unterschrieben), erfüllt nach unserem Dafürhalten ein solches System keine der relevanten Transparenzregularien (siehe unten).

Ein wesentlicher Sicherheitsaspekt neben dem privaten Schlüssel stellt in diesem Szenario auch der Ort der Speicherung der signierten Ergebnisse dar, speziell, mit welchen Privilegien auf diese Daten zugegriffen und mit welchen Privilegien sie verändert werden können.

Eine Möglichkeit einer technischen Lösung des Problems des nachträglichen Einfügens von Audit-Daten stellt die Einführung einer Hash-Chain dar (Schneier & Kelsey, (1999)). Dies bedeutet, dass die einzelnen Log-Einträge mit Hilfe einer kryptographischen Hash-Funktion miteinander verknüpft werden. Diese Verknüpfung kann bspw. durch eine andere Entität geschehen, als die Trusted Entity, welche die Logs signiert, wodurch ein zusätzlicher Layer an Sicherheit erreicht wird. Auf weitere Varianten der Speicherung und Ablage des Logs, speziell im Bezug darauf, wo der Transparency log gespeichert werden soll, wird im nächsten Unterkapitel

noch genauer eingegangen. Basierend auf dieser Methode existieren einige Abarten, die versuchen die Abhängigkeit von physischen Entitäten zu vermindern, indem bspw. die Kettenbildung direkt vom Datenspeicher, bspw. einer Datenbank, durchgeführt wird. Grundsätzlich kann natürlich ein Angreifer, der die volle Kontrolle über das System erhält, jederzeit ab dem Zeitpunkt der Kontrolle die weiteren Logs manipulieren. Allerdings kann sichergestellt werden, dass vergangene Logs nicht mehr verändert werden können. Dies ist natürlich hinfällig, sollte es zu einer Korrumpierung aller beteiligten Entitäten kommen. Dieser Fall ist irrelevant, da auch alle Kontrollinstanzen korrumpiert wären, bzw. die Angreifer einfach das System neu aufsetzen und von vorne starten könnten. Eine Abart dieses Konzepts basiert auf der Absicherung des Transaction Logs gegenüber Manipulationen durch das verwandte Konzept der Chained Witnesses. Dieses Konzept speichert eine verkettete und signierte Version des Transaction Logs als Audit Trail und kann dadurch nicht nur seine Lückenlosigkeit garantieren, sondern protokolliert auch gleichzeitig alle Transaktionen mit (siehe Abbildung 18). Darauf aufbauend können auch Audit & Control-Mechanismen gebaut werden, die bspw. einen Doctor-in-the-Loop-Ansatz (Kombination aus Expertensystem und menschlichem Arzt für die Diagnose, Human-in-the-Loop-Ansatz im medizinischen Bereich) für beide Seiten gegenüber Manipulationen absichern (Kieseberg et. al., (2016)).

*Abbildung 2020 – Chained Transaction Log*




In diesem Konzept können bereits alle Transaktionen eindeutig nachvollzogen werden, allerdings steht dieses Konzept in direktem Widerspruch zum Recht auf Vergessenwerden, da die Nachvollziehbarkeit nur solange gewährleistet ist, als keinerlei Schritte aus dem Audit & Control entfernt werden. Werden zusätzlich die Anreicherungsinformationstabellen in die Datenbank einbezogen, so könnten Daten auch wiederverarbeitet werden. Im Fall von Anreicherungstabellen sollte prinzipiell darauf geachtet werden, dass Datenfelder nicht gelöscht oder verändert werden, sondern nur ihr jeweiliger Gültigkeitsbereich neu definiert wird, sodass eine Nachbearbeitung jederzeit möglich ist. Auch diese oftmals durch Regularien geforderte Funktionalität kann mit dem Recht auf Vergessenwerden kollidieren, dies ist allerdings ein allgemeines Problem, dass ein nachvollziehbares Auditing bestenfalls Pseudonymisierung zum Schutz der Privatsphäre zulässt.

Wesentlich bei der Durchsetzung dieses Aspekts der DSGV ist nach unserer Meinung besonders die Frage, welches Recht und welche Pflicht als höherwertiger anzusehen sind: Das Recht auf Vergessenwerden, oder die lückenlose Nachvollziehbarkeit, bzw. sogar eine etwaige Forderung des Reprocessings. Hierbei wird es nach unserem Dafürhalten keine allgemeingültige Entscheidung geben, sondern eine, die auf den jeweiligen Use-Case und die Art der Verarbeitung abstellt.

Transparenz im Zusammenhang mit der Verarbeitung von personenbezogenen Daten kann auch insofern eine Hürde darstellen, da keine Standard-Schemata für personenbezogene Daten bzw.

noch keine allgemein anerkannten „Best Practices“ für entsprechende Granularität der Transparenz-aufzeichnungen existieren. Im Bereich der Forschung aus dem Bereich des Semantischen Web existieren einige Vorschläge zu Ontologien zur Beschreibung von persönlicher Daten und deren Provenienz, vgl. z.B. (Bartolini et al. (2015)). Jedoch handelt es sich bei dieser und ähnlichen Arbeiten mehr um akademische Schemata denn um Standards. Im Bereich der Standardisierung eignet sich etwa das vom W3C propagierte PROV Vokabular[11] für die Beschreibung von Datenprozessierungsschritten, es bietet jedoch kein konkretes Vokabular zur Beschreibung von personenbezogenen Daten, etwa um aus den Transparenz-Aufzeichnungen automatisiert Fragen wie die Folgenden beantworten zu können:

- Wer bzw. welche Institution oder juristische Person hat Daten über ein Individuum erfasst?
- Welche Daten wurden erfasst oder welche personenbezogenen Merkmale sind aus diesen Daten ersichtlich? z.B.:
    - Standort oder Bewegungsmuster einer Person in einem zeitlichen Kontext
    - Die Kommunikation oder Beziehungen zwischen Personen
    - Bilder oder Tonaufzeichnungen, die eine Person identifizieren
    - Andere identifizierende Merkmale wie Sozialversicherungsnummer, Kontonummer, etc.
    - Behandlungen oder Dienste die von einer Person in Anspruch genommen wurden
- Standardisierte Verwendungszwecke

Momentan existieren keinerlei umfassende Schemata oder Vokabulare um all diese Aspekte in ausreichend klarer und austauschbarer Art und Weise zu beschreiben, sodass die Erfüllung der Transparenzpflichten lediglich von Fall zu Fall festgestellt werden kann. Wir nehmen daher an, dass die Entwicklung und Einführung solcher Standards einen positiven Effekt zur leichteren Verarbeitbarkeit und Überprüfung von Transparenzaufzeichungen hätte. Im Folgenden gehen wir auf technische Aspekte der Realisierung einer Transparenzschicht, die entsprechende Aufzeichnungen speichert, genauer ein.

### 5.4.1 Technische Realisierung der Transparenzschicht

Neben Aspekten des Widerspruchs mit dem Recht auf Löschung bzw. Vergessenwerden, gibt es verschiedene Ansätze um eine Transparenz-Schicht, welche die Verarbeitung personenbezogener Daten praktisch aufzeichnet, zu implementieren.

Die unterschiedlichen Ansätze haben verschiedene Vor- und Nachteile: Um den Transparenz-Verpflichtungen (in Bezug auf den Nachweis der Verarbeitung von personenbezogenen Daten, aber auch dem Nachweis der Benutzer-Einwilligung zur Speicherung und Verarbeitung von Daten (‚consent’), oder der Aufforderung von Berichtigung oder Löschung) im Sinne der DSGV nachzukommen, obliegt dem Datenverarbeiter eine Reihe von Aufzeichnungspflichten diverser Daten-Transaktionen: Wer hat Daten an wen weitergegeben? Zu welchem Verwendungszweck

[11] PROV,https://www.w3.org/TR/prov-overview/

und unter welchen konkreten Bedingungen? Wie wurden die Daten verarbeitet, anonymisiert, und aggregiert? Ziel dieses Abschnitts ist es, bestehende Lösungen zur Implementierung einer Transparenzschicht zu untersuchen und bezüglich Ihrer Stärken und Schwächen in Bezug auf bestimmte geforderte Funktionen zu vergleichen, bzw. generelle technische Herausforderungen, die noch nicht gelöst sind, zusammenzufassen. Um den Pflichten zur Transparenz nachzukommen und dies auch in einer auditierbaren Art und Weise nachweisen zu können werden die folgenden Kernfunktionen in einer Transparenzschicht benötigt.

#### 5.4.1.1 Anforderungen an die Transparenzschicht

Bevor wir konkret auf verschiedene Logging Architekturen ("Ledgers") eingehen, definieren wir im Folgenden eine Liste von technischen und nicht-technischen Anforderungen.

***Ledger Funktionalität***:

- Vollständigkeit (Completeness): Jegliche Ereignisse im Zusammenhang mit Datenverarbeitung und -weitergabe sollen im Ledger aufgezeichnet werden.
- Vertraulichkeit (Confidentiality): Die datenverarbeitende Organisation und die betroffene Person (data subject) sind die einzigen, die Transaktionen betreffend Ihrer Daten am Ledger einsehen können.
- Korrektheit (Correctness): Die Aufzeichnungen am Ledger sollen die Verarbeitung korrekt wiedergeben.
- Unveränderbarkeit (Immutability): Die log Informationen am Ledger müssen unveränderbar und persistent sein, sodass eine nachträgliche Änderung oder Neudefinition der Transaktions-Historie nicht möglich ist, d.h. dass die Log-Historie als gesamtes vor Veränderungen geschützt sein muss.
- Integrität (Integrity): Zusätzlich müssen die einzelnen log Informationen am Ledger müssen vor versehentlicher oder beabsichtigter Modifikation geschützt sein.
- Interoperabilität (Interoperability): Die Log-Informationen am Ledger sollen über Organisationsgrenzen hinweg integrierbar sein, in dem Sinne, dass es den betroffenen Personen möglich sein soll, log Informationen, die sie von verschiedenen Datenverarbeitern bekommt, zu kombinieren und zu beurteilen.
- Unleugbarkeit (Non-repudiation): Es soll nicht möglich sein, Ereignisse im Zusammenhang mit Datenverarbeitung und –weitergabe, welche auf dem log festgehalten wurden, im Nachhinein zu leugnen.
- Richtigstellung und Löschung (Rectification & Erasure): Es soll Betroffenen möglich sein, Fehler in personenbezogenen Daten beheben zu lassen bzw. auf Anfrage der Betroffenen deren Löschung zu veranlassen.
- Nachverfolgbarkeit/Nachvollziehbarkeit (Traceability): Es soll möglich sein die einzelnen Schritte der Datenverarbeitung nachzuverfolgen. Demzufolge soll es möglich sein, Ereignisse in Bezug auf die (Weiter-)Verarbeitung von personenbezogenen Daten, in einer Art und Weise zu verbinden, die eine solche Nachverfolgung ermöglicht.

**Ledger Robustheit:**

- Verfügbarkeit (Availability): es muss, unabhängig davon ob Ledger-Informationen lokal oder verteilt gespeichert werden, sichergestellt sein, dass keine Informationen verloren gehen bzw. diese jederzeit zugänglich sind. Insbesondere hat dieser Aspekt starke Verbindungen zur Datensicherheit, da eventuelle Sicherheitslücken oder das Schließen von Sicherheitslücken die Verfügbarkeit nicht beeinträchtigen dürfen.

- Performanz: die implementierte Lösung muss entsprechenden, dem Use-Case angepassten, Datendurchsatz erlauben und Ereignisdaten schnell genug verarbeiten können. Optimierungen wie parallel-Verarbeitung oder spezielle Indexstrukturen sollten benutzt werden, um die erforderliche Effizienz sicherzustellen.

- Skalierbarkeit (Scalability): Der Ledger muss mit entsprechend großen Datenvolumina bei den Ereignisdaten umgehen können.

- Speicherung (Storage): Um die Menge an direkt auf dem Ledger gespeicherten Informationen gering zu halten, sollten Ereignisdaten selbst getrennt gespeichert werden und lediglich ein Hashwert der Ereignisdaten sowie ein Zeiger zu den tatsächlichen Daten am Ledger selbst vorhanden sein.

#### 5.4.1.2 Der Status Quo

Um Log-Einträge zur Provenienz von Ereignisdaten im Sinne der Transparenz persistent zu speichern existieren drei generelle Möglichkeiten, die sich gegenseitig nicht vollständig ausschließen: jeder Datenverarbeiter verwaltet eine lokalen Transparenzschicht (Local Ledger), welche zusätzlich remote als Backup (etwa von den Betroffenen) gesichert werden kann; eine global verwaltete Transparenzschicht (Global Ledger) wird von einer von sowohl Datenverarbeiter als auch Betroffenen vertrauenswürdig erachteten Drittorganisation verwaltet (trusted third party - TTP); oder eine global verwaltete Transparenzschicht (Global Ledger) wird verteilt in einer Peer-to-Peer Architektur (P2P) gespeichert. (Bonatti et. al., (2017)) fassen mögliche Transparenzschicht-Architekturen unter diesen Gesichtspunkten zusammen.

##### 5.4.1.2.1 Beispiele für Local Ledger Architekturen

Als erste Möglichkeit kann jeder Datenverarbeiter Ereignisse und dazugehörige Provenienz-Informationen, inklusive Ereignisse zur Datenweitergabe (weitergegebene sowie erhaltene Daten von Datensubjekten) lokal speichern. Ein Nachteil gegenüber der alternativen Speicherung bei einem vertrauenswürdigen Dritten (auch TTP - trusted third party) könnte hier die Sicherstellung der Wiederherstellbarkeit von Ereignisdaten sein, wenn der physische Rechner, auf dem die die Log-Informationen gespeichert wurden kompromittiert wurde, da bei lokaler Speicherung der Datenverarbeiter selbst für die Sicherheit und Verfügbarkeit der Daten sorgen muss. Bellare und Yee (Bellare & Yee, (1997)) sowie Schneier und Kelsey (Schneier & Kelsey, (1998)) haben jeweils Ansätze präsentiert, um ein Verschlüsselungs- und Signaturschema basierend auf sogenannten Message Authentication Codes (MACs) im Zusammenhang mit Hashing-Algorithmen zu realisieren. Dieses kann benutzt werden, um Sequenzen (chains) von log Einträgen zu erstellen, die herangezogen werden, um Vertraulichkeit und Integrität zu garantieren. MACs (deutsch: Nachrichtenauthentifizierungscode) dienen dazu, den Ursprung von Daten oder Nachrichten auf Integrität überprüfbar zu machen und sind selbst symmetrische Schlüssel die über kollisionsresistente kryptographische Hash-Funktionen generiert und verifiziert werden. Diese können laut (Bellare & Yee, (1997)) benutzt werden um (i) die

Vertrauenswürdigkeit von Log-Informationen; (ii) die Unveränderbarkeit von Log-Einträgen, sowie (iii) die Erkennung von Löschungen von Log-Einträgen sicherzustellen. Der Basis-MAC Schlüssel, welcher benötigt wird, um die Integrität des Logs zu garantieren muss hier allerdings durch eine TTP zur Verfügung gestellt werden. (Schneier & Kelsey, (1998)) benutzen ebenfalls MACs, allerdings ist in Ihrem Falle das Log selbst eine Sequenz von Hashes (hash chain), anstatt der verschlüsselten log Ereignisse (cipher blocks). Holt schlägt eine weitere Alternative vor in der Public Key Cryptography mit Hash Chains kombiniert wird (Holt, (2006)). Dieser Ansatz wurde von Ma und Tsudik weiter verbessert, die zeigen wie individuelle log-Eintrags Signaturen in einer einzigen aggregierten Signatur kombiniert werden können, welche dazu benutzt werden kann, die einzelnen Komponenten-Signaturen zu verifizieren und das Log vor Abschneiden (truncation) zu schützen (Ma & Tsudik, (2009)). Sackmann et al. wenden entsprechende Logging-Methoden konkret auf Datenschutz-Szenarien und präsentieren ein System für Privacy-aware Logging (Sackmann et. al., (2006)). Zusätzlich, führen Sie das Konzept des Datenschutzbeweises ("privacy evidence") ein und diskutieren hier wie entsprechende Log-Einträge benutzt werden können, um Datenschutzstrategien von Benutzern zu verifizieren.

In Bezug auf Robustheit, evaluieren sowohl Bellare and Yee, als auch Holt Performance und Skalierbarkeit Ihrer Logging und Verifikations-Algorithmen. Auch Ma and Tsudik bieten einen Vergleich alternativer Signaturgenerierungs und Verifikations-Algorithmen mit existierenden Methoden. Allerdings ist in keiner dieser Evaluierungen der tatsächliche Aufwand in der Praxis in Bezug auf einen bestimmten Use-Case und in Bezug auf realistisches Datenaufkommen in Logging-Szenarien aus der heutigen Big Data Praxis erwähnt. Daher muss die Frage, ob diese Methoden in der Praxis anwendbar sind unbeantwortet bleiben.

#### 5.4.1.2.2 Beispiele für Global Ledger Architekturen mit Trusted Third Party

Die zweite Gruppe von Architekturen aus der Literatur benutzt eine zentral von einer oder mehreren TTPs verwaltete Transparenzschicht. Accorsi benutzt hier wiederum MAC-basierte sichere Logging-Verfahren, die hier zum Logging von Daten ressourcenbeschränkter Geräte benutzt werden können, um die Daten remote zu loggen (Accorsi (2006)). Wouters et al. betonen, dass Daten oft zwischen verschiedenen Prozessen ausgetauscht werden, sodass entsprechende Logging-Ereignisse nicht isoliert betrachtet werden können, und damit die Notwendigkeit besteht, die Aufzeichnungen von Ereignissen zentral bzw. integriert zu speichern (Wouters et. al., (2008)). Die Autoren verwenden wiederum Public Key Cryptography um Ereignisse zu loggen, sodass Datensubjekte den Process-Status verifizieren können. (Hedbom et. al., (2009)), (Peeters et. al., (2013)) und (Pulls et. al., (2013)) präsentieren ähnliche Logging Mechanismen die eine zentrale Transparenzschicht für Datensubjekte zur Verfügung zu stellen. In allen diesen Fällen wird ein Protokoll basierend auf sicheren MAC Logging Mechanismen benutzt, um Vertraulichkeit sowie Unlinkability (d.h. es ist nicht möglich Ereignisse zu verschiedenen Datensubjekten und von verschiedenen Datenverarbeitern am Log zu verlinken) sicherzustellen. Außerdem kann das zentrale Log jeweils über mehrere Server verteilt werden. Im Falle von Peeters et al., (2013) sowie Pulls et al., (2013), ist jeweils jeder log Eintrag aus einem User-Block, einem Datenverarbeiter-Block und den verschlüsselten Ereignis-Daten zusammengesetzt. Eine TTP stellt die MAC-Schlüssel zur Verfügung, bzw. nimmt die Verschlüsselung mit dem öffentlichen Schlüssel der Benutzer vor, signiert die Ereigniseinträge weiters mit dem eigenen privaten Schlüssel und setzt die Datensubjekte von neuen Einträgen in Kenntnis, analog wird der Block für den Datenverarbeiter (data processor block) generiert. Die Log Informationen und personenbezogene Daten werden so verschlüsselt, dass jeweils nur das Datensubjekt und der Datenverarbeiter zu relevanten Einträgen Zugang haben. Im Falle von Datenweitergabe wird ein neuer abgeleiteter öffentlicher Schlüssel generiert, sodass der Private Schlüssel des

Datensubjekts jegliche mit diesem abgeleiteten öffentlichen Schlüssel verschlüsselte Informationen entschlüsseln kann). Dieser abgeleitete Schlüssel, welcher vom zweiten Datenverarbeiter benutzt werden kann, kann auch dazu benutzt werden um sicherzustellen dass Logdaten zu Einzelpersonen nicht einfach von unberechtigten Dritten verknüpft werden können (Unlinkability von Logs): es gibt also mehrere unterschiedliche, abgeleitete öffentliche Schlüssel, zu jedem privaten Schlüssel (also pro Datenverarbeiter zu Daten eines individuellen Datensubjekts und umgekehrt).

Peeters et al., (2013) und Pulls et al., (2013) beschreiben jeweils Evaluierung der Performance ihrer Algorithmen im Hinblick auf Datendurchsatz von einer lokalen und globalen Perspektive. Die Autoren betonen hier, dass speziell Verschlüsselung und Signatur die zeitkritischen Operationen sind, und dadurch die Zeit der Erstellung von log-Einträgen nicht linear mit der Größe der log-Einträge skalieren kann. Ebenso sind Verifikations- und Entschlüsselungsschritte in ähnlicher Weise mit erheblichem Rechenaufwand verbunden und können dadurch einen Engpass (gerade im Hinblick auf Big Data Szenarien) darstellen.

#### 5.4.1.2.3 Beispiele für Global Ledger Architekturen mit Peer-to-Peer Netzwerk

In einem verteilten Architektur-Ansatz werden Log-Informationen (in einem “virtual global Ledger”) über ein Peer-to-Peer-Netzwerk verteilt, wobei Einträge auf jedem Peer repliziert werden. Schneier and Kelsey (siehe oben) betonen speziell die Risiken im Zusammenhang mit einer einzigen TTP und diskutieren solche Architekturen in Bezug auf Möglichkeiten wie *n* nicht notwendigerweise vertrauenswürdige Maschinen dazu verwendet werden können eine einzige TTP zu ersetzen, wobei eine verteilte Anzahl von *m* verteilten Maschinen benötigt wird, um die Basis-MAC Schlüssel zu erstellen. Ein Hauptvorteil einer solche Architektur liegt darin, dass durch die Replikation sowohl Verfügbarkeit als auch Redundanz und damit Datensicherheit der Log-Einträge sichergestellt werden kann. Weitzner et al. diskutieren weiters wie Transparenz und Nachweisbarkeit/Verantwortlichkeit (accountability) gegenüber einer bestimmten Policy in einer verteilten Peer-to-peer Architektur unter Verwendung von existierenden Standard Web-Protokollen sichergestellt werden kann (Weitzner et. al., (2008)). Sogenannte “accountability peers” sind in so einer Architektur um die Einhaltung von vereinbarten Datenzugangspolicies zu prüfen, entsprechende Ereignislogs zur Aufzeichnung von Audit-Trails zu verwalten und die Verifikation (etwa anhand von Methoden des automatischen Schließens über formal beschriebene Policies) von Fragen zur Nachweisbarkeit/Verantwortlichkeit der Policy-Einhaltung zu ermöglichen. Leider beschränkt sich der Artikel weitgehend auf eine Beschreibung der Anforderungen, anstatt eine konkrete Architektur vorzustellen. Seneviratne und Kagal entwickeln aufbauend auf diesen Ideen Vorschläge wie in einem verteilten Netzwerk von Peers (d.h. Knoten in einem Peer-to-Peer-Netzwerk) verschlüsselte Transaktions-Daten gespeichert werden können (Seneviratne & Kagal, (2014)). Der Fokus der Arbeit liegt hier aber weniger in der verteilten Architektur und deren Funktionalität/Charakteristiken selbst, sondern mehr in Methoden zur Verifikation der Einhaltung von User-Policies in Bezug auf Verwendungseinschränkungen von Daten aber auch dem Aufzeigen/Erklären von aus Aktionen/Ereignissen abgeleiteten Informationen und deren Implikationen auf den Datenschutz.

Eine weitere alternative Architektur von Zyskind et al. verwendet die in letzter Zeit (v.a. durch Crypto-Currencies) populäre Blockchain-Technologie, um Zugang zu personenbezogenen Daten zu managen und zu loggen (Zyskind et. al., (2015)). Die Blockchain-Technologie basiert per se auf peer-to-peer Netzwerken und Verschlüsselung. Die Autoren beschreiben wie das Blockchain Daten-Modell und Application Programming Interfaces (APIs, Schnittstellen zu Funktionalitäten für andere Anwendungen) erweitert und angepasst werden kann, um sowohl Datenzugriffe als

auch Verarbeitungs-Transkationen aufzuzeichnen. Daten werden hier mittels eines verteilten Schlüssels verschlüsselt, an die Blockchain gesendet, wobei die Ereignisdaten selbst abseits der Blockchain (in einem, lokalen, oder von einer TTP verwalteten, key-value store) gespeichert werden und nur ein Zeiger auf die Daten in Form eines Hash auf dem Public Ledger gespeichert wird. “Compound identities” (Zusammengesetzte Identitäten) werden benutzt um sicherzustellen, dass nur die Betroffenen Datensubjekte und Datenverarbeiter/Service Provider, die Zugang zu den Daten haben sollen die Daten entschlüsseln können. Die Adaptierung der Blockchain-Technologie um Zugangskontrolle und transparente Speicherung von Transaktionen durchführen zu können, steht bei dieser Arbeit im Vordergrund, wobei die tatsächliche Eignung einer solchen Architektur (etwa im Kontext der anderen vorgeschlagenen Möglichkeiten) für die Realisierung einer Transparenzschicht zur Datenverarbeitung nicht im Zentrum steht. Bei allem Hype um Blockchain-Technologien sollte man sich hier im Klaren sein, dass viele “Features” von Blockchain nicht per se für die Realisierung von Transparenz-Logs entwickelt wurden, sondern viele der Eigenschaften von Blockchain zum anonymen Zahlungsverkehr etwa für die Implementierung einer solchen Transparenzschicht gar nicht notwendig sind, sowie mit beträchtlichem Aufwand verbunden sind. Auch bestehen im Zusammenhang mit Big Data Anwendungen beträchtliche berechtigte Zweifel im Zusammenhang mit der Skalierbarkeit von bestehenden Blockchain Implementierungen.

In der Literatur sind Robustheitsaspekte von P2P Architekturen im Vergleich zu Architekturen, die eine TTP benutzen noch kaum untersucht worden. Eine genaue Analyse der nicht-funktionalen Aspekte von P2P-Ledgers oder Blockchains als Basis für eine Transparenzschicht ist daher momentan noch nicht möglich. Es sei erwähnt, dass speziell in Blockchains verbreitete Voting-techniken im P2P Bereich, Manipulationen ermöglichen, wenn eine Organisation mehr als die Hälfte aller Peers kontrolliert. Dies ist speziell bei privaten Blockchains bei geringer Anzahl von Peers zu beachten.

#### 5.4.1.3 Lückenanalyse im Vergleich der betrachteten Architekturen

Im Vergleich der betrachteten Architektur-Optionen kristallisieren sich folgende Haupt-Herausforderungen heraus, die sich hauptsächlich aus Konflikten zwischen bestimmten in den oben aufgelisteten Anforderungen ergeben (Bonatti et. al., (2017)).

***Correctness, Completeness & Non-Repudiation:*** Korrektheit und Vollständigkeit können, unabhängig von der Architektur nicht innerhalb des Systems garantiert werden, da über eine Transparenzschicht allein nicht sichergestellt werden kann, dass nicht von Vornherein inkorrekte oder unvollständige Aufzeichnungen auf dem Ledger gelogged werden. Sogenannte „Fair exchange“-Protokolle können potentiell genutzt werden um Non-repudiation (Nichtabstreitbarkeit) zu einem gewissen Grad zu garantieren (d.h., es wird per Protokoll und entsprechender Einträge sichergestellt, dass keine der beteiligten Parteien im Nachhinein abstreiten kann, dass ein bestimmtes Ereignis oder eine bestimmte Transaktion stattgefunden hat), diese Mechanismen wurden aber bisher noch nicht im Zusammenhang mit Transparenz Logging Mechanismen im Kontext der Datenverarbeitung verwendet, es bestehen also keine Standardlösungen.

***Confidentiality & Integrity:*** Die Kombination von MAC zusammen mit entweder der Speicherung von verschlüsselten Ereignisdaten selbst oder entsprechenden Hash-Chains scheint momentan der Standard-Ansatz zu sein, um Vertrauenswürdigkeit und Integrität von Logs sicherzustellen, benötigen jedoch eine TTP zur Vergabe der Basis MACs. Obwohl Schneier and Kelsey beschreiben, dass es möglich sein sollte diese TTP durch *n* nicht notwendigerweise vertrauenswürdige Maschinen in einem P2P Netzwerk zu ersetzen von denen ein Quorum von

*m* benötigt wird um einen Basis MAC secret key zu erstellen, werden keine Details oder konkrete Implementierungen dieses Ansatzes beschrieben. Zusätzlich sind in unserem Szenario Mechanismen zur Berichtigung und Löschung von personenbezogenen Daten nötig, ohne dabei die Integrität des Logs zu zerstören, was ein weitgehend ungelöstes Problem darstellt.

***Immutability, Rectification & Erasure***: Die DSGV schreibt vor, dass es möglich sein muss Daten zu berichtigen oder zu löschen (auf Initiative der Datensubjekte, etwa unter dem "Recht auf Vergessenwerden"), gleichzeitig soll es aber den Datenverarbeitern nicht möglich sein, die log Informationen neu zu erfinden oder zu ändern. Das Recht auf Vergessenwerden kann hier als physische Löschung verstanden werden, wo Daten sowohl aus dem System als auch vom Log entfernt werden müssen. Wie bereits besprochen zählen in diesem Zusammenhang sowohl die Sicherstellung der Gesamtintegrität des Logs, als auch die Sicherstellung, dass die Daten im System tatsächlich vollständig gelöscht wurden (Korrektheit & Vollständigkeit) Hauptherausforderungen dar. Eine mögliche Lösung wäre hier kryptographische Löschung (Zerstörung des Schlüssels, anstatt Löschung der verschlüsselten Daten, d.h. die Daten wurden verschlüsselt, werden selbst nur als freier Speicherplatz markiert, lediglich der Schlüssel wird unwiederbringlich zerstört, damit sind die Daten (theoretisch) unwiederbringlich gelöscht) und Update via Versionskontrolle (die allerdings im potentiellen Konflikt zur Vollständigen Löschung stehen, außer wenn sie entsprechend mit kryptographischen Löschmethoden kombiniert werden.

***Interoperability & Traceability:*** Ein weiterer kritischer Punkt ist die Interoperabilität über mehrere dezentrale Logs hinweg. Nachdem die Forschung und Systeme zu Logging-Mechanismen sich bisher zumeist auf bestimmte Betriebssystem- oder Transaktionsereignisse in Anwendungen (wie geschlossene Datenbanksysteme) fokussiert haben sind Interoperabilitätsaspekte von Logs bisher weniger behandelt worden. Obwohl es im Bereich der Nachverfolgbarkeit von Log-Informationen (traceability) einige Arbeiten gibt, beschränken sich diese zumeist auf die Verknüpfung von zusammenhängenden Ereignissen innerhalb eines einzelnen Logs. In unserem Zusammenhang, wo Datensubjekte im Sinne der Nachvollziehbarkeit der Verarbeitung und Weitergabe ihrer personenbezogenen Informationen die Möglichkeit haben sollten, Log Informationen von mehreren Datenprovidern zu integrieren stellen sich hier zusätzliche Herausforderungen. Momentan existieren keinerlei umfassende Standards oder akzeptierte Best-Practices im Sinne von zu verwendenden Formaten, Schemata oder Vokabularen, um alle relevanten Aspekte bezüglich des Austauschs und der Verarbeitung von personenbezogenen Daten in klarer und austauschbarer Art und Weise zu beschreiben.

***Performance & Scalability:*** Nachdem die geänderten Rahmenbedingungen im Datenschutz vor allem auf Big Data-Anwendungen zurückzuführen sind, wo mit hohen Datenvolumina auch und vor allem in der Speicherung von personenbezogenen Daten zu rechnen ist, stellt ein feingranuläres Logging von entsprechenden Ereignissen bezüglich der Sammlung, Verarbeitung und Weitergabe solcher Daten in einer Transparenzschicht eine hohe technische Herausforderung an die Skalierbarkeit bestehender Logging-Mechanismen dar. Die Granularität dieser Aufzeichnungen ist ein Aspekt, der nicht zuletzt auch im Sinne der rechtlichen Interpretation der Aufzeichnungspflichten zu verstehen ist, wobei zu erwarten ist, dass sich die technischen Möglichkeiten und damit auch die Machbarkeit der Aufzeichnung auch auf feinerer Granularität stetig weiterentwickeln. Kosten und auch Energie/Nachhaltigkeit von Datenspeicherung, und notwendigen Datenverarbeitungsschritten im Zusammenhang mit Encryption und Hashing bei komplexem Logging sind ein weiterer kritischer Aspekt, der

zunehmend an Bedeutung gewinnt.[12] In der Verarbeitung von Ereignisdaten werden daher Optimierungen wie etwa Parallelverarbeitung, und/oder Indexierungsmethoden relevant sein. Des Weiteren können Komprimierungsverfahren eine Rolle spielen um den Datentransfer zu optimieren. Die Optimierung von Abfragen und (Replikation von) Updates von wachsenden Logs in verteilten Datenbasen stellen eine weitere Herausforderung in Bezug auf Performance und Skalierbarkeit dar.

***Storage*:** Die Speicherung aller Log-Informationen in einem einzigen Log-Server oder Replikation auf alle Peers in einem P2P-Netzwerk ist im Allgemeinen nicht skalierbar möglich, da theoretische Herausforderungen in verteilten Systemen wie das CAP-Theorem (das darstellt, dass Konsistenz, Verfügbarkeit, und Ausfalltoleranz in einem verteilten System niemals gleichzeitig erreichbar sind) auch im Zusammenhang mit Transparenz-Logs zu erwarten sind (Seth & Lynch, (2002)). Eine Möglichkeit ist es nun, Log-Daten modular zu verteilen bzw. zu replizieren in verteilten, modularen Ledgers, mit verteilten TTPs und hierarchisch verbundenen P2P-Netzwerken.[13] Allerdings, stellt so eine Architektur neue Herausforderungen an die Fehlertoleranz. Die Relevanz der gespeicherten Informationen und die richtige Wahl der Granularität, sowie "careful forgetting" könnte hier wie schon im Zusammenhang mit Skalierungsaspekten erwähnt helfen, um Storage-Anforderungen und –Kosten zu reduzieren, indem nur relevante Information gespeichert wird, um die Einhaltung bestimmter Policies (Rahmenvorschriften für Berechtigungen und Verbote) zu prüfen, die in einer bestimmten Anwendungsdomäne nötig sind, und irrelevante Ereignisse (periodisch) zu löschen oder nur selektiv, in Abhängigkeit bestimmter Policies zu speichern.

***Availability:*** Im Sinne der Verfügbarkeit ist es nötig, entsprechende Redundanz und Sicherheitsmechanismen in der Transparenzschicht sowie regelmäßige Backup-Mechanismen vorzusehen, wobei entsprechende Best-Practices am Stand der Technik einzuhalten sind, um den/die log Server zu schützen. Hier könnten etwa öffentlich verfügbare Blockchains eine Rolle spielen, um die Wiederherstellbarkeit zu ermöglichen, indem anstatt zentral oder bei einer einzelnen TTP zusätzlich hashes der log Information auf einer öffentlichen Blockchain (wie Bitcoin) gespeichert werden. Allerdings geben die meisten dieser öffentlich verfügbaren Blockchains im Moment keinerlei Service Level Agreements (SLAs), im Gegensatz zu TTPs.

### 5.4.1.4 Vorschläge für mögliche Lösungen

Zuallererst fehlt ein standardisiertes, maschinenlesbares Austausch- und Repräsentationsformat für Transparenzinformationen und Policies.

Als Standardformat eignet sich etwa das „Resource Description Framework" (RDF), ein Standard-Metadaten Format aus dem Bereich Semantic Web, das es ermöglicht Informationen austauschbar und dereferenzierbar am Web als „Linked Data" zu speichern und auszutauschen. Dieses Format könnte es ermöglichen, Log-Informationen und Ereignisse zu repräsentieren und untereinander zu verlinken, sowie diese in standardisierter, maschinen- und menschenlesbarer Form verfügbar zu machen. In einer vom BMVIT mitfinanzierten Studie wurde unlängst das Potenzial dieser Technologien, die ursprünglich aus dem Bereich der Web-Standards kommen, im Enterprise Bereich beleuchtet (Fernández et. al., (2016)). Speziell bietet RDF die Möglichkeit, Standard-Vokabulare und Metadaten-Standards zu integrieren um Datenprovenienz oder die

[12] https://www.forbes.com/sites/peterdetwiler/2016/07/21/mining-bitcoins-is-a-surprisingly-energy-intensive-endeavor/#994315a5bbf0

[13] ähnlich wie bei "Sharding", ein Begriff aus der NoSQL-Welt, wo verteilte Daten nicht vollständig auf allen Knoten repliziert werden, sondern kategorisch verteilt werden, um etwa zu garantieren, dass in einer Cloud, Daten möglichst nah am Benutzer liegen.

Beschreibung von Policies in semantisch eindeutiger Art und Weise maschinenlesbar zu beschreiben. Damit könnten momentan offene Interoperabilitätsaspekte adressiert werden, wobei die Verlinkung zwischen Ledgers über Linked-Data-Prinzipien zusätzlich Traceabiliy-Aspekte über Ledgers hinweg adressieren könnte und damit die Entwicklung darauf basierender automatisierter Compliance-Checking-Mechanismen ermöglicht. Eine Reihe relevanter Vokabulare und Standards existieren in diesem Zusammenhang bereits. Die PROV-[14] und OWL-Time[15]-Ontologien eignen sich beispielsweise um Datenprovenienz und temporale Aspekte von Log-Informationen zu repräsentieren. Erweiterung des PROV-Modell speziell zur Modellierung der Verarbeitung personenbezogener Daten wären hier allerdings notwendig. OWL-Time bzw. die Synchronisation von temporalen Aspekten ist speziell für verteilte Log-Informationen relevant. Beispielsweise wenn Audit-Trails über verschiedene Systeme und Ledgers verteilt sind, ist die Synchronisation von Zeitstempeln und die Sicherstellung der richtigen Reihenfolge kritisch.

Um in standardisierter Art und Weise repräsentierte Transparenz-Informationen verarbeiten zu können, werden weiters effiziente, skalierbare Methoden zur Abfrage und Integration dieser Daten benötigt. Hier spielen Abfragemechanismen und semantische Integration temporaler Information eine wichtige Rolle; diese Aspekte sind allerdings zum Teil im Bereich RDF und Semantic Web noch mit offenen Forschungsfragen verbunden. Unter verschiedenen Vorschlägen zu temporalen Erweiterungen von RDF und Abfragesprachen zu temporalen Archiv-Informationen, hat sich bisher noch kein Standard durchgesetzt (Fernández Garcia et. al., (2016)). Zusätzlich existieren verschiedene generelle Ontologien im Bereich der Modellierung genereller Event-Daten wie etwa die Event[16]- und LODE[17]-Ontologien die potenziell erweitert oder adaptiert werden könnten, um Datenverarbeitungs-Events zu modellieren (Rinne et. al., (2013)). Ein zusätzlicher Vorteil von Linked-Data ist die Verfügbarkeit von Ontologien zur Beschreibung von Datenverarbeitungs-Policies und Forschungsergebnissen im Zusammenhang mit Verifikationsmechanismen zur Einhaltung solcher Policies. Die Skalierbarkeit solcher Methoden stellt allerdings noch eine offene Forschungsfrage dar. Als ein Beispiel für Arbeiten in diesem Zusammenhang seien sog. „Knowledge Compilation"-Ansätze erwähnt die semantischen Metadaten in eine kompakte Policy-Repräsentation kompilieren die effizient verarbeitet und verifiziert werden kann (siehe Ansätze basierend auf partieller Evaluation dieser Policies in (Bonatti et. al., (2002)).

Die Nutzung von RDF und URIs zur eindeutigen Identifizierung von Policies und Verbindung mit Standard Web Protokollen könnte weiters dazu benutzt werden verteilte Ledgers miteinander zu verbinden, im Gegensatz zur Verwendung eines einzelnen monolithischen oder auch zu einem bestimmten P2P-Ledger. In diesem Zusammenhang arbeitet etwa auch das W3C an Standard-Protokollen um verteilte Ledgers modular zu verbinden (siehe auch das „Interledger protocol" (Hope-Bailie & Thomas, (2016)).

#### 5.4.1.5 myData

Wie diese Ausführungen zeigen, ist die Entwicklung geeigneter Lösungen im Bereich Transparenz und Nachvollziehbarkeit mit hohen Investitionskosten verbunden und daher auf kleinere und mittelständische Unternehmen nicht ohne weiteres übertragbar. Eine Win-Win Situation in der europäischen und österreichischen Wirtschaft könnte dadurch herbeigeführt werden, dass entsprechende offene Standard-Lösungen in Zusammenarbeit zwischen Firmen, der öffentlichen

[14] PROV,https://www.w3.org/TR/prov-overview/
[15] OWL-Time,https://www.w3.org/TR/owl-time/
[16] http://motools.sourceforge.net/event/event.html
[17] http://linkedevents.org/ontology/

Verwaltung und der universitären und privaten Forschung im Rahmen von Public Private Partnerships entwickelt werden.

Ein Vorzeige-Beispiel in diesem Zusammenhang ist die hauptsächlich finnisch-estnische MyData Initiative (mydata.org), der sich aber inzwischen bereits etliche andere EU- und nicht EU-Länder (z.b. auch Japan, Schweiz) angeschlossen haben.[18] Mydata hat zum Ziel offene technische Lösungen und Prinzipien zum Umgang mit personenbezogenen Daten zu definieren und zu entwickeln. Grundidee des Ansatzes ist, dass die AnwenderInnen mittels offenener Softwarelösungen ihre Berechtigungen selbst zentral verwalten können, wobei die entsprechenden Daten von „Trusted Entities" (und nicht von vielen einzelnen Diensteanbietern) verwaltet werden.

Mydata und wird in Ländern wie Finnland und Estland von sowohl Einrichtungen der öffentlichen Verwaltung sowie Firmen und VertreterInnen der Zivilgesellschaft mitgetragen. Die zweite von MyData organisierte Konferenz wurde unlängst mit über 700 internationale TeilnehmerInnen in Helsinki und Tallin abgehalten. Hauptthemen, die auch in der Umsetzung der DS-GVO und damit verbundener technischer Lösungen begründet sind waren etwa Personal Clouds, Identity Management, dezentrale Verwaltung von Transparenz- und Consent-Aufzeichnungen, sowie semantische Interoperabilität von Lösungen zum Datenschutz und zur Verwaltung Transparenz-Aufzeichnungen und der Dokumentation von Einverständniserklärungen (unter dem Label „consent/privacy interoperability") etc.,

Das Label „**My Data"** kann hier als Spiegelbild der (in Österreich bereits sehr aktiven) „**Open Data** „Initiativen gesehen werden, wo es darum geht, einerseits nicht-sensitive Daten der öffentlichen Verwaltung (aber immer mehr auch von Firmen) unter freien Lizenzen wiederbenutzbar und offen zur Verfügung zu stellen (Open Data) und andererseits personenbezogene Daten die an verschiedenen Stellen der öffentlichen Verwaltung benötigt werden oder gesammelt werden über eine einheitliche technisch transparente Plattform, transparent und kontrollierbar für die BürgerInnen und Bürger zur Verfügung zu stellen (My Data). Ziel einer derartigen Public Private Partnership ist, dass sich letztendlich auch Firmen an diese Plattform andocken können, um unter Verwendung derselben offenen Lösungen den regulativen Anforderungen an Transparenz und Datensicherheit um Umgang mit personenbezogenen Daten nachzukommen. Die Verwendung von offenen Plattformen ermöglicht hier einerseits Synergien, aber auch die Entwicklung neuer Geschäftsmodelle für Startups etwa im Bereich Personal Clouds, Identity Management, dezentrale Verwaltung von Transparenz-Aufzeichnungen und Consent, etc. Die Einbeziehung der Zivilgesellschaft in entsprechende Initiativen ist insofern dringend notwendig, um eine breite Vertrauensbasis der Bürgerinnen und Bürger in entsprechende Lösungen und Standards sicherzustellen.

## 5.5 Ausgewählte Methoden zur Absicherung von Informationen

Grundsätzlich liegt einer der wesentlichen Aspekte der neuen datenzentrierten Businessmodelle in der Auslagerung von Datenanalysen an entsprechende Experten. Diese Einbindung ist natürlich datenschutzrechtlich hochbrisant und relevant und die Daten müssen im Allgemeinen, so eine Auswertung durch Externe nicht durch die Beauftragung abgedeckt ist, entsprechend anonymisiert werden. Allerdings besitzen auch die anonymisierten Daten oftmals ein sehr hohes Schutzpotential. Dies liegt nicht nur in rechtlichen Hintergründen begründet, sondern vielmehr

[18] https://mydata.org/hubs/ letzter Zugriff 9.9.201792017

darin, dass die entsprechenden Daten oftmals einen hohen Wert für das Unternehmen besitzen. Daher ist der tiefergehende Schutz dieser Daten oftmals aus anderen Erwägungen heraus extrem, relevant Dabei wurden im Rahmen des Projekts die folgenden Ansätze diskutiert:

### 5.5.1 Generierung synthetischer Daten

Grundsätzlich kann es für manche Fragestellungen ausreichen, gar nicht mit Echtdaten zu hantieren, sondern lediglich mit Daten, die die gleichen Eigenschaften besitzen. Die Lösung kann daher in einigen Fällen in der Erstellung sogenannter Synthetischer Daten liegen, das sind Datensätze, die in gewisser Hinsicht echten Datensätzen (bis auf eine genau abschätzbare und vordefinierte Abweichung) gleichen, dennoch keinerlei Rückschlüsse auf die originalen Daten ermöglichen. Je nach Fragestellungen kann die Nutzung solcher synthetischen Daten ausreichend sein und bietet den zusätzlichen Vorteil, dass sie keinerlei Schutzmechanismen und Anforderungen der DSVGO unterliegen, da es sich hierbei um rein „fiktive“ Daten handelt, die nicht sensibel sein können. Speziell im Bereich des Testens von Algorithmen ist eine derartige Lösung oftmals ausreichend, da es lediglich um das Feststellen der Performance eines Algorithmus in Hinblick auf Daten einer gewissen Gestalt, nicht jedoch um die tatsächlichen Analyseergebnisse geht.

Zur Generierung von synthetischen Daten stehen verschiedene Werkzeuge und Ansätze zur Verfügung, die als Input entweder auf abstrakten Datenmodellen (bspw. statistischen Eigenschaften), oder aber auf Echtdaten beruhen. Speziell in letzterem Fall, oftmals als semi-synthetische Daten bezeichnet, muss darauf geachtet werden, dass nicht zu viel Information von den Echtdaten in die synthetischen Daten fließt und so u.U. Rückschlüsse auf sensible Informationen ermöglicht werden (Skopik et. al., (2014)).

Bekannte Werkzeuge zur Generierung von synthetischen Daten umfassen bspw. den IBM Quest Synthetic Data Generator[19], den DTM Dat Generator, Loresoft[20] und ist auch weiterhin Inhalt akademischer Forschungsarbeiten (Singla et. al., (2016)).

### 5.5.2 Weitergehende Anonymisierung und Löschung

Grundsätzlich kann in vielen Fällen soweit anonymisiert werden, dass die relevanten Informationen gänzlich aus den Daten verschwinden. Dazu können einzelne Datensätze entfernt, Nutzdaten gelöscht, oder gar falsche Informationen hinzugefügt werden, um den Datensatz für andere wertlos zu machen. Allerdings ist ein derartig gesicherter Datensatz oftmals auch für weitere, erwünschte, Analysen unbrauchbar und ein Austausch damit hinfällig.

Dieser Ansatz ist daher typischerweise nicht zielführend und kann in den meisten Fällen verworfen werden.

### 5.5.3 Aufbau eines Research Servers

Die Bereitstellung einer abgesicherten Analyseumgebung, wie sie bspw. im DEXHELPP Research Server (DRS) erfolgt[21], erlaubt die gemeinsame Bearbeitung von Daten durch unterschiedliche BearbeiterInnen in einer speziell gesicherten und abgetrennten Umgebung, die zusätzlich zu den Daten auch alle benötigten Analysewerkzeuge zur Verfügung stellt, somit eine

[19] https://sourceforge.net/projects/ibmquestdatagen/
[20] http://www.loresoft.com/DataGenerator-Generate-Intelligent-and-Realistic-Test-Data
[21] http://www.dexhelpp.at/

Extraktion und Weitergabe sensibel Daten hinfällig macht. Zusätzlich werden den BearbeiterInnen die Daten lediglich in der notwendigen Granularität zur Verfügung gestellt, bzw. wird gar kein direkter Zugriff auf die Informationen benötigt, sondern es werden lediglich analytische Ergebnisse zurückgeliefert. Zusätzlich zu dieser Abschottung wird das System sehr stark überwacht: Alle Abfragen, die durch die BearbeiterInnen getätigt werden, sowie alle Algorithmen und Programme werden eingehend geloggt und regelmäßig kontrolliert. Dabei werden auch automatisierte Ansätze eingesetzt, die Missbrauch rasch aufdecken können.

In Hinblick auf den DEXHELPP Research Server befindet sich die gesamte Infrastruktur innerhalb eines virtuellen privaten Netzwerks (VPN), auf das nur die Mitglieder des DEXHELPP Konsortiums Zugriff haben. Das VPN wurde durch eine dedizierte Hardwarelösung realisiert und entspricht aktuellen Industriestandards, ist stark verschlüsselt und gewährt nur jenen Mitgliedern Zugriff, die sich anhand eines sicheren Passwortes sowie eines Tokens eindeutig authentifizieren können. Durch diese Zwei-Faktor-Authentifizierung mittels des Passworts und eines zeitabhängigen Schlüssels der nur vom Token (dem zugeordneten Smartphone des DEXHELPP Mitglieds) generiert werden kann, wird ein sehr hohes Sicherheitsniveau erreicht, der dem aktuellen Stand der Sicherheitstechnik entspricht.

Ein ausgeklügeltes Berechtigungssystem stellt sicher, dass nur jene Projektmitglieder Zugriff auf Daten- und Rechenressourcen haben, die eine Freigabe für das konkrete Projekt besitzen. Diese Freigaben werden dynamisch, d.h. je nach den Anforderungen des aktuellen Projektverlaufs angepasst, was bedeutet, dass Berechtigungen sowohl dynamisch hinzugefügt, als auch wieder entzogen werden können. Diese Zuordnung von Berechtigungen erfolgt nachvollziehbar und transparent.

Der Datenaustausch findet prinzipiell und ausschließlich über die verschlüsselte und abhörsichere VPN-Leitung statt. Das VPN Netzwerk selbst ist vom allgemeinen Internet durch eine Firewall abgetrennt und erlaubt keinen Internetzugriff direkt von den einzelnen virtuellen Maschinen des DRS. Einzige Ausnahme ist die sogenannte Datenschleuse, die den einzigen Punkt darstellt, über den Daten von außerhalb in die DRS Infrastruktur gelangen und auch wieder verlassen können. Diese Datenaustauschoperationen erfolgen innerhalb eines eigenen Systems, das nur auf Anforderung hin und nach der automatischen Überprüfung der Berechtigungen die Austauschdienste zur Verfügung stellt.

Optional können Datensets durch digitale Wasserzeichen individuellen Nutzern zugeordnet werden, wodurch etwaige Datenlecks direkt Personen zugeordnet werden können, die mit dem Datenset gearbeitet haben. Zusätzliche Anonymisierungsmaßnahmen verhindern den Abfluss sensibler Daten, die einzelnen Personen zugeordnet werden könnten.

### 5.5.4 Fingerprinting zur Erkennung von Datenweitergabe

Fingerprinting und Watermarking sind zwei Techniken, die es ermöglichen Daten zu markieren. Ursprünglich oftmals zum Schutz von Multimediadaten entwickelt (Langelaar et. al., (2000)) und (Li et. al., (2005)), sind sie auch im Bereich der Weitergabe strukturierter Daten zunehmend wichtig geworden. Der wesentliche Unterschied ist der, dass Watermarking-Techniken lediglich versuchen, eine Markierung zum Ursprung der Daten einzubringen, d.h. entweder zu zeigen, dass die Daten echt sind, oder dass sie von einer bestimmten Quelle erzeugt wurden (und bspw. dieser auch gehören). Solche Watermarks basieren oftmals auf Techniken, die sogenannte „Markierungsdatensätze“ in die Datenmenge einfügen, d.h. es werden zusätzliche synthetische Datensätze erzeugt, die als Markierungen dienen. Dabei muss darauf geachtet werden, dass

diese Markierungsdatensätze nicht zu viel Einfluss auf die statistischen Eigenschaften des Datensets haben und damit nicht zu viel Beeinflussung durch die Markierung eingebracht wird. Umgekehrt, basieren manche Mechanismen auf der Entfernung von Datensätzen aus der Datenmenge.

Fingerprints hingegen zielen darauf ab, eine bestimmte Kopie eines Datensets einer bestimmten Ressource zuzuweisen (Liu et. al., (2004)). Dies ist sinnvoll, wenn Daten im Rahmen eines kollaborativen Vorhabens an mehrere Partner weitergegeben werden müssen. Werden diese bspw. im Internet angetroffen, so ist nicht nur interessant, ob es sich tatsächlich um die eigenen Daten handelt, sondern auch, welcher Partner die Daten weitergegeben hatte. Auch hierbei gibt es Ansätze basierend auf Markierungsdaten, wobei vor allem die einfachen Ansätze sehr anfällig gegenüber kollaborativen Angriffen sind, d.h. Angriffe, bei denen mehrere Datenrezipienten zusammenarbeiten, bspw. um Markierungsdaten herauszurechnen. Auch in dieser Hinsicht gibt es einige Verfahren, die mit kollaborierenden Angreifern umgehen, oder aber auch andere Angreifermodelle mitigieren können (Gross-Amblard, (2013); Sion et. al., (2002)).

Ein weiterer Ansatz basiert direkt auf Basis der k-Anonymisierung selbst (Kieseberg et. al., (2014)). Dieser Ansatz hat den entscheidenden Vorteil, dass, falls die Anonymisierung aus datenschutzrechtlichen Gründen sowieso durchgeführt werden muss, der Fingerprint quasi intrinsisch gleichzeitig und mehr oder weniger von selbst erzeugt wird. Zusätzlich bietet dieses Verfahren den Vorteil, dass die implizierte Verzerrung klar ersichtlich ist und damit keinerlei vorgetäuschte Genauigkeit entsteht.

Die Idee liegt darin, dass jeder Kollaborationspartner die gleichen Daten in unterschiedlicher Form anonymisiert bekommt, d.h. die Grunddatenauswahl unterscheidet sich nicht, der Unterschied liegt lediglich darin begründet, welche Quasi-Identifier in welcher Granularität vorhanden sind. Die Erkennung von Datenlecks ist daher denkbar einfach:

- Vor der Weitergabe, im Fall des Beispiels an zwei Partner U1 und U2, wurden verschiedene Generalisierungsstrategien zur Anonymisierung gewählt.
- Die gewählten Anonymisierungsstrategien werden gespeichert.
- Wird ein Datensatz angetroffen, so ist es sehr einfach, die Anonymisierungsstrategie aus der Gestalt des Datensatzes selbst zu errechnen. Diese Strategie wird mit der Liste der gespeicherten User abgeglichen und das Datenleck kann erkannt werden.

## 5.6 Forschungsfragen zu den Themen Anonymisierung und Datenlöschung

Basierend auf den in diesem Projekt durchgeführten Analysen haben sich die folgenden Forschungsfragen als wesentlich herauskristallisiert. Dabei handelt es sich nicht nur um rein technische Fragestellungen, sondern vor allem um Fragen, die einen integrierten Forschungsansatz zwischen technischen und rechtswissenschaftlichen Experten erfordern:

- Die Wahl der konkreten Sicherheitsparameter zur Anonymisierung, speziell des Sicherheitsfaktors „k“ im Rahmen von k-anonymity oder verwandten Verfahren. Das gleiche gilt auch für den Einsatz von Differential Privacy, hier ist die Wahl des Faktors Epsilon zu klären. Im Fall von Datenperturbation, d.h. der Verschneidung von Echtdatensätzen mit synthetischen Daten, ist zu klären, ab welchem Verhältnis zwischen Echtdaten und Perturbationsdaten die Privacy der beteiligten Personen gewahrt bleibt.

- Speziell im Fall von Sensordaten kann die Einteilung, ob es sich bei den Daten um sensible Informationen handelt, nicht-trivial sein. Hier wird, unter Umständen branchenspezifisch, zu klären sein, wodurch sich Quasi Identifier auszeichnen und generelle Kriterien festzulegen, wie diese zu erkennen und mit ihnen umzugehen ist.

- Entstehen im Rahmen interner Datenverarbeitung sensitive Datenströme durch die Verschneidung von (u.U. teilweise sensiblen) Daten, so wäre zu klären, ab wann eine Anonymisierung durchgeführt werden muss, bzw. ob der Akt der Verschneidung noch unanonymisiert durchgeführt werden darf. Dies ist speziell wichtig, da eine Verschneidung anonymisierter Daten oftmals nicht möglich ist.

- Welche Form des Löschens ist ausreichend und welche forensischen Methoden existieren? Dies umfasst auch die Entwicklung neuer forensischer Methoden, die einfach umzusetzen sind und speziell in sehr komplexen Systemen vorhandene, nicht gelöschte, Metainformationen zur Datenrekonstruktion ausnutzen. Dies ist extrem relevant, um den durch die DSGVO implizierten Schutz der Daten durch Löschen auch in der Realität umzusetzen. Dabei ist die Frage nicht nur auf „physikalisches“ oder „logisches“ Löschen beschränkt, sondern umfasst auch den Umgang mit Backups, Sicherheitsmechanismen, internen (Security-)Logs, sowie anderen Methoden fortschrittlichen Datenmanagements.

## 5.7 Einverständniserklärungen („Informed Consent“)

Möglichkeiten zum Einholen expliziter Einwilligung/Zustimmung zur Verwendung oder Weitergabe von personenbezogenen Daten von betroffenen Personen („informed consent“) sind ein zentraler Bestandteil der DSGV. Diese Zustimmung kann dabei auch elektronisch eingeholt werden. Dazu muss der Verwendungszweck aus diesen Einwilligungen klar hervorgehen („purpose specification“) und die Verarbeitung personenbezogener Daten hat nicht über diesen Verwendungszweck hinauszugehen (‘use limitation’). Im folgenden Kapitel gehen wir näher auf damit verbundenen technischen Herausforderungen ein, diese Art der Einwilligung technisch zu unterstützen.

Ein Hauptproblem für Firmen in diesem Zusammenhang besteht hier darin, dass die Verwendung und Analyse von personenbezogenen Daten zu neuen Analysen und zum Anbieten neuer Dienste trotz der Verfügbarkeit großer entsprechender Datenmengen nicht ohne weiteres möglich ist, da entsprechende Einwilligungen für neue, ursprünglich nicht vorgesehen Verwendungszwecke erst eingeholt werden müssen. Vorgefertigte, in statischen Endnutzerverträgen festgelegte, Nutzungsbedingungen reichen nicht aus. Es besteht der Bedarf nach technologischen Methoden einerseits von Seiten der Datenverarbeiters dynamisch solche Einwilligungen einzuholen, aber auch von Seiten der Datensubjekte dynamische die Einwilligung zu widerrufen oder anpassen zu können. Weiters soll es Betroffenen insbesondere technisch ermöglicht werden, inkorrekte Daten, die sie betreffen, zu korrigieren bzw. deren Korrektur zeitnah zu verlangen.

### 5.7.1 Anforderungen an Einverständniserklärungen

Wir definieren im Folgenden eine Reihe von Hauptaspekten im Zusammenhang mit Einverständniserklärungen und diskutieren den Stand der Technik in Bezug auf bestehende Technologien und die technische Umsetzbarkeit der einzelnen Aspekte.

**Funktionalität:**

- *Kategorisierung (Categorisation)*: Um den Benutzer nicht zu überfordern oder mit unnötig vielen Interaktionsschritten zu konfrontieren, sollte es möglich sein, Anfragen um Einwilligung von Verarbeitungsschritten zu gruppieren und per Kategorie nach Einwilligung zu bitten, ohne damit die Eindeutigkeit des Verwendungszwecks zu unterminieren.

- *Anpassung (Customisation)*: Im Gegensatz zu derzeit üblichen „Alles oder Nichts"-Ansätzen die generellen Nutzungsbedingungen und Einwilligungen der Verarbeitung personenbezogenen Daten an eine monolithische Einwilligungserklärung knüpfen, sind angepasste technische Lösungen, die es Datensubjekten ermöglichen, Einwilligungen partiell und feingranular zu kontrollieren wünschenswert. Damit soll festgelegt werden, welche Daten unter welchen Bedingungen verarbeitet bzw. weitergegeben werden können.

- *Innovationsfreiheit (Innovation Readyness)*: Von Seiten der Datenverarbeiter ist es wünschenswert, dass die Einwilligung zur Datenverarbeitung soweit wie möglich auch die Entwicklung neuer innovativer Dienste und Business-Intelligence-Lösungen ermöglicht, ohne jedes Mal den Benutzer mit neuen detaillierten Einwilligungsanfragen konfrontieren zu müssen.

- *Nachträgliches Einverständnis (Retrospective Consent)*: Ein weiterer Aspekt für datenverarbeitende Firmen ist die Unantastbarkeit von Daten für neue Verwendungszwecke für deren Verwendung noch keine explizite Einwilligung existiert, und die Umsetzung technischer Möglichkeiten solche Einwilligung möglichst unkompliziert (für beide, den Datenverarbeiter und die BenutzerInnen) einzuholen.

- *Entziehung des Einverständnisses (Revocation)*: Den Datensubjekten muss jederzeit die Möglichkeit eingeräumt werden, ihr Einverständnis zur Verarbeitung von Daten zur Gänze oder teilweise zurückzuziehen.

- *Verständlichkeit (Understandability)*: Einverständniserklärungen/-anfragen müssen in verständlicher Form präsentiert werden und in einer Art und Weise, die die BenutzerInnen nicht überfordert. Dies ist speziell in Big Data-Szenarien herausfordernd, wo komplexe Datenverarbeitungsschritte stattfinden, die auf nicht-trivialen Algorithmen und Verarbeitungsschritten basieren, die den BenutzernInnen nicht leicht näherzubringen sind. Es ist hier einerseits sicherzustellen, dass die BenutzerInnen verstehen, worauf sie sich durch die Zustimmung einlassen, als auch zu vermeiden, den/die BenutzerIn mit den Details der Verarbeitung zu überfordern. Des Weiteren spielt eine zugängliche Präsentation der Einverständnisanfrage (User-Interface-Aspekte) eine Rolle.

**Robustheit:**

- *Leistungsfähigkeit (Performance):* Wie schon bei den Aspekten rund um Transparenz soll die Verarbeitung der Einholung oder die Verifikation von Einverständnis (etwa durch automatisches überprüfen von user policies) keine signifikanten Performance-Einbußen bedeuten. Auch hier können, wie im Kapitel zu Transparenz beschrieben, Parallelverarbeitung oder Indexing-Methoden zur schnelleren Verarbeitung und gesteigerten Effizienz Anwendung finden.

- *Skalierbarkeit (Scalability):* Personalisierte Policies und eine vollständige Prüfung der Einhaltung anhand der Daten der Transparenzschicht in Echtzeit stellen eine signifikante Herausforderung an die Skalierbarkeit eines Systems dar, dass das Einverständnis automatisch prüfen soll.

- *Speicherung (Storage):* Um personalisierte Policies kompakt speichern und überprüfen zu können, wäre eine Möglichkeit etwa nur Abweichungen von Standard-Policies und personalisierte Änderungen von Policies explizit zu speichern, generell besteht die Anforderung hier zwischen den Speicheranforderungen für die Policies selbst und der Skalierbarkeit und Leistungsanforderungen an deren Verarbeitung abzuwägen, was vom konkreten Use-Case abhängig sein kann, ähnlich bei der Granularität der Kontrollmöglichkeiten.

## 5.7.2 Wissenschaftlicher Status Quo

Im Rahmen der akademischen Forschung wurden etliche Fragestellungen zur Einholung von Einverständnis, bzw. zu Visualisierungsmethoden im Zusammenhang mit Einverständnis oder der Verwaltung von Zugriffsrechten wissenschaftlich untersucht, wobei es hier einige noch offene Fragestellungen gibt.

### 5.7.2.1 Methoden zur Einholung von Einverständnis

Generell existieren drei Methoden zur Einholung von Einverständnis: i) Generelle Nutzungsbedingungen (etwa in Form eines Vertrags); ii) Opt-In; und iii) Opt-Out, wobei letztere beiden Methoden auch elektronisch erfolgen können.

Im Moment benutzen Firmen hauptsächlich natürlichsprachige Policies und Beschreibungen der Nutzungsbedingungen, denen der Benutzer entweder elektronisch in der Form impliziter genereller Nutzungsbedingungen zustimmt oder diese bei Aktivierung des Dienstes aktiv annimmt/unterschreibt. Implizite Nutzungsbedingungen sind nach der DSGV keine valide Form der Einwilligung, da die Einwilligung nicht explizit und freiwillig gegeben wurde, falls der Benutzer keine andere Wahl hat als die Einwilligung zu geben. ein Hauptproblem von umfangreichen Policy-Dokumenten ist weiters, dass diese oft nicht flexibel genug sind um personalisierte Datenverarbeitungspolicies zu ermöglichen, etwa um Mehrwertdienste an bestimmte partielle Zustimmungen zu knüpfen oder die partielle Rücknahme von Einverständnis zu erlauben.

"Opt-In" ermöglicht es, explizite Einwilligung von BenutzerInnen für bestimmte Services einzuholen, z.B. durch affirmative Unterschrift oder mittels Online-Formular, unter Sicherstellung der Identität/Authentifizierung. Allerdings zeigt die Forschung, dass Benutzer typischerweise kaum Zeit darauf verwenden den Text solcher Online-Formulare zu lesen und zu verstehen, bevor die Zustimmung „per Click“ gegeben wird. Es ist daher wissenschaftlich höchst umstritten, ob Opt-in-Modelle die Kriterien informierter, expliziter Einwilligung erfüllen. McDonald und Cranor beschreiben folgende Abschätzung: "[we] estimate that reading privacy policies carries costs in time of approximately 201 hours a year, worth about $3,534 annually per American Internet user. Nationally, if Americans were to read online privacy policies word-for-word, we estimate the value of time lost as about $781 billion annually.” (McDonald et. al., (2008))

Im Gegensatz dazu wird, "Opt-Out" primär im medizinischen Bereich verwendet, wo BenutzerInnen in generellen Bedingungen zustimmen, dass Ihre Daten für Forschungszwecke verwendet werden können, außer wenn sie explizit (opt-out) eine Bekundung abgeben von der Teilnahme exkludiert zu werden. Dieser Opt-Out-Ansatz wird etwa in biomedizinischer Forschung

verwendet, wenn eine Studie ein niedriges Risiko im Sinne des Datenschutzes darstellt oder wenn Opt-In- Einverständnis nicht machbar ist oder das Studien-Design beeinträchtigen würde. Nach Vayena et al. fordern Forscher eine breitere Akzeptanz dieses Opt-Out-Ansatzes für Forschungszwecke, und argumentieren damit, dass Forschung & Entwicklung (F&E) bzw. Innovation damit ermöglicht wird, ohne die Autonomie der BenutzerInnen zu verletzen (Vayena et. al., (2013)). Wenn die BenutzerInnen eine positive Aktion ausführen muss (etwa „Clicken", um ein Formular zu bestätigen), und der einholende Datenverarbeiter eine klar ersichtliche Nachricht an prominenter Stelle anbringt, etwa die Bestätigung, dass eine „Opt-Out-Box" NICHT gewählt wurde, reicht dies nach Meinung der Autoren als Einwilligung aus.

Die Frage, ob Opt-In die einzig zulässige Methode zum Einholen von Einwilligung sei, ist also nicht eindeutig zu beantworten.

#### 5.7.2.2 Anpassung von Consent

Benutzeroberflächen spielen bei der Anpassung von Einverständnis bzw. der Kontrolle von erlaubten Verarbeitungen sowie bei der Verständlichkeit eine kritische Rolle. Ein Beispiel ist das Zugriffsrechtesystem des mobilen Betriebssystems Android für Anwendungen von Drittanbietern: Eine Studie, die die Benutzer-Aufmerksamkeit, deren Verhalten und Verständlichkeit des Systems untersucht hat legt dar, dass die TeilnehmerInnen oft von Zugriffsrechten gar keine Notiz nehmen und dass nur wenige (17%) während der Installation von Anwendungen darauf achten (Felt et. al., (2012)). RequestPolicy[22], ein Werkzeug zur Steigerung des Datenschutzes in Webbrowsern durch Kontrolle von sog. cross-site requests (Samuel & Zhang, (2009)), schlägt Whitelists (Inklusionslisten) vor, um die Privatsphäre der User zu schützen. Blacklists (Ausschlusslisten) werden als unzureichend erachtet, nachdem sie neue Anfragen, von bisher unbekannten Seiten nicht behandeln können. Google bietet eine Lösung zur Visualisierung und teilweisen oder gänzlichen Löschung von personenbezogenen Daten. Nach Login in Google's Dashboard können verschiedene gespeicherte Daten in verschiedene Kategorien unterteilt eingesehen werden, z.B. Suchhistorie, Kalender, Kontakte, etc. Zusätzlich, kann die sog. „location history", die geographischen Positionsdaten mit Zeitverlauf, (typischerweise über Android-Smartphones erfasst), auf einer Karte visualisiert und (teilweise oder zur Gänze) gelöscht werden. Nachdem die Datenmengen oft unüberschaubar sind, erlaubt das Interface die Erstellung von Zusammenfassungen und Aggregate von Daten wo möglich. Des Weiteren kann die Erfassung verschiedener Daten gestoppt werden. Was die Definition von Policies betrifft, gehen einige existierende anwenderfreundliche Schemata über Zugangsbeschränkungen und Zugriffsrechte hinaus: So bietet etwa Creative Commons (CC) ein Schema für Datennutzungslizenz-Templates welche kombinierbar sind und mit sprechenden Akronymen und Piktogrammen (Icons) hinterlegt sind (z.B. "BY" für die Verpflichtung zur Attribuierung/Namensnennung, "$" für kommerzielle Weiterverwendung, etc.). Allerdings existiert momentan kein solches akzeptiertes Schema für Endkunden-Policies zu personenbezogenen Daten, deren Verarbeitung und Weitergabe. Ein rudimentärer aber nicht sehr ausgegorener Ansatz, der im Beta-Stadium verblieben ist, stellen Mozillas „Privacy Icons" dar.[23]

#### 5.7.2.3 Einschränkungen von Einverständnis

(Acquisti et. al., (2013)) sowie (Borgesius & Zuiderveen, (2015)) verweisen auf etliche jüngere Studien die Zweifel an der Effektivität von Transparenz-Benachrichtigungen und Einwilligungs-Auswahl zur Erklärung von Einverständnis äußern. Die Studie von Acquisti et al. etwa berichtet,

[22] https://www.requestpolicy.com/
[23] https://wiki.mozilla.org/Privacy_Icons

dass ein Mehr an Kontrolle zu fahrlässiger Handhabung und mehr Risiko führte, eine Beobachtung die sie als “control paradox” bezeichnen. Zusätzlich fanden die Autoren heraus dass sogar kurze Pausen zwischen der Präsentation der von Datenschutzerklärungen/-bedingungen durch dazwischen präsentierten irrelevanten Informationen genügen um die Aufmerksamkeit für bzw. das Verständnis von Datenschutzerklärungen/-bedingungen zu vermindern oder gar zu eliminieren. Nach Borgesius tendieren Menschen dazu, solchen Bedingungen zuzustimmen oft ohne sich im Klaren zu sein welche Ihrer Daten nachverfolgt werden, sodass ihre Einwilligung nicht als Einverständniserklärung zu werten ist. Borgesius streicht weiters eine starke Tendenz unter Usern hervor, Standardeinstellungen zu akzeptieren und kurzfristigen Wert durch einen Dienst vor längerfristigen Risiken in den Vordergrund zu stellen.

### 5.7.3 Lückenanalyse

Aufgrund der Literatur lassen sich etliche offene Forschungsfelder und–fragen ableiten.

***Categorisation & Understandability:*** Die geringe Aufmerksamkeit, und das geringe Verständnis, das BenutzerInnen Einwilligungserklärungen zu personenbezogenen Daten laut Felt et al., Acquisti et al. und Borgesius entgegenbringen, legen nahe, dass um die Benutzbarkeit von Mechanismen zur Einverständniserklärung zu verbessern, die Benutzerinteraktion verändert/verbessert werden muss, um BenutzerInnen die effektive Verwaltung von Zugriffs- und Weitergaberechten in einer verständlichen Form zu ermöglichen. Für jede Organisation, die über Daten einer Person verfügt, sollten die Zugriffsrechte und Bedingungen in einheitlicher verständlicher Form visualisiert und kategorisiert abrufbar sein. Diese Kategorien sollten in einer Art und Weise angezeigt werden, die eine feingranulare Verwaltung und Modifikation der Zugriffsrechte erlaubt. BenutzerInnen sollten mit spezifischen Anfragen von Firmen konfrontiert werden, welche den Verwendungszweck und die dafür notwendigen Daten klar darlegen, möglichst mit leicht verständlichen Beispielen und gleichwertigen Optionen für den/die BenutzerIn zuzustimmen oder abzulehnen. Wo möglich können hier die Konzepte von hierarchischen Datenschutz-Policies wie von der Art. 29 Working Party[24] propagiert eingesetzt werden, die leicht verständliche Icons oder verständliche Kurztexte komplementär zu leicht zugänglichen Detailinformationen vorschlägt um mehr Klarheit und Transparenz zu erreichen.

***Customisation, Retrospective Consent & Revocation:*** Zur Formalisierung von Policies in maschinenlesbarer Form existieren verschiedene Policy Sprachen wie XACML[25], ODRL (McRoberts & Rodríguez Doncel, (2015)), KAoS (Bradshaw et. al., (1997)), Rei (Kagal & Finin, (2003)) oder Protune (Bonatti & Olmedilla, (2007)), welche dazu benutzt werden können personalisierte Zugangs-, Verwendungs- und Weitergabe-Policies zu beschreiben und dynamisch Anpassung und (partielle) Rücknahme von Einverständnis zur Verwendung von personenbezogenen Daten zu ermöglichen. Eine Herausforderung hier stellt die Abbildung von menschenlesbaren Policies in solch maschinenlesbarer Form dar, aber auch die verifizierbare Prüfung durch haltbare Beweise, dass die Einhaltung durch die maschinenlesbaren Policies und entsprechende Algorithmen zu deren Durchsetzung garantiert ist, sowie die Erstellung von Werkzeugen zu deren einfacher Verwaltung und zur Annahme durch die Industrie. Arbeiten von Villata et al. (Villata & Gandon, (2012)) oder die Ergebnisse des EU Forschungsprojekts PrimeLife

[24] Article 29 Data Protection Working Party (2004), Opinion 10/2004 on More Harmonised Information Provisions: Adopted on 25th November 2004. 11987/04/EN. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp100_en.pdf.

25 XACML v3.0 Privacy Policy Profile (2010). Available at: http://docs.oasisopen.org/xacml/3.0/xacml-3.0-privacy-v1-spec-cd-03-en.pdf

können hier als mögliche Startpunkte dienen wo Technologien zur Erstellung und Formalisierung modularer, Benutzer-verständlicher Policy-Vorlagen entwickelt wurden, um 'CreativeCommons-ähnliche', leicht merkbare Schemata für Endnutzer-Policies durch Akronyme und Piktogramme mit maschinenlesbarer Semantik zu erstellen (Holtz et. al., (2011)).

***Innovation:*** Ein Hauptproblem im elektrischen Einholen von Einverständnis zur Datenverarbeitung ist, dass Organisationen neue Geschäftsmodelle und –Möglichkeiten oft erst durch Analyse von personenbezogenen Daten entdecken können, um etwa Kundeninteressen zu erforschen. Es existiert hier gewissermaßen ein Henne-Ei Problem indem Firmen die Einwilligung der Betroffenen benötigen, um Benutzerdaten zu analysieren bevor sie den genauen Verwendungszweck überhaupt kennen, also nicht angeben können wofür sie überhaupt eine Einwilligung benötigen. Borgesius stellt weiters hervor, dass Firmen (die Investition in) Technologien zur Sicherung der Privatsphäre selten als mögliche Chance für einen kompetitiven Vorteil erkennen. Gürses et al. (Gürses et. al., (2016)) sowie Hansen (Hansen, (2016)) bieten hier Anleitungen zu „Privacy Engineering", wie es in der Praxis angewandt werden kann und stellen ebenfalls nicht-technische Herausforderungen dar: diese Herausforderungen betreffen Benutzererfahrung im Umgang mit den benutzten Systemen, das mangelnde Bewusstsein von Themen um den Datenschutz in Ingenieurs-Teams und des damit verbundenen Vergessens von Datenschutz im System-Design, sowie innerhalb der Prioritäten der Organisation. Datenschutz im Design (privacy-by-design) und Datenschutz als Default (privacy-by-default) müssen also schon beim Systemdesign mitgedacht werden, was alle Stakeholder betrifft und damit immer noch eine offene Frage darstellt.

### 5.7.4 Mögliche Lösungsansätze

Anstatt monolithischer, statischer Einverständniserklärungen besteht eine Notwendigkeit Technologien zu entwickeln, die die dynamische, automatisierte Anpassung von von elektronisch erteilten Einwilligungen erlauben, mit speziellem Fokus auf rechtliche und auch ethische Aspekte sowie Benutzbarkeit. Solche Technologien und Mechanismen sollten es einerseits BenutzerInnen ermöglichen Daten zu korrigieren, existierende Policies zu Ihren Daten einzusehen und anzupassen. Wie schon im Abschnitt zu Transparenztechnologien erwähnt, existieren auch hier mögliche Ansätze aus dem Bereich Linked-Data: Kirrane et al. (Kirrane et. al., (2017)) geben in ihrem Survey einen Überblick zu Techniken der Zugriffskontrolle für RDF, die etwa benötigt würden, um Linked Data als Paradigma zum Austausch sensitiver, durch Zugangs-Policies geschützter Daten zu verwenden.

Außerdem sind etliche der beschriebenen maschinenlesbaren Policy-Sprachen bereits in RDF modelliert, die zur Beschreibung und Verifikation von Nutzungs-Policies, rechtlichen Regularien und Businessregeln verwendet werden können und mit Datenprovenienz-Informationen und Transparenzinformationen und –Ereignissen verknüpft werden können zur automatischen Verifikation von Endnutzer-Policies. Allerdings existieren hier im Sinne der Skalierbarkeit auch offene Forschungsfragen zur notwendigen Ausdrucksstärke solcher Sprachen und zur Komplexität und Skalierbarkeit entsprechender Verifikationsmethoden. Vokabulare zur Beschreibung von Policies wie ODRL, die derzeit von der Permissions and Obligations working group des W3C standardisiert werden leiden nach wie vor unter teilweiser semantischer Ambiguität die einer Maschinenverarbeitung im Weg stehen, bzw. könnten sie sich in der Praxis als unvollständig oder unzureichend entpuppen um komplexe Policies zur Verarbeitung personenbezogener Daten zu beschreiben (Steyskal & Polleres, (2015)). Zusätzlich muss RDF (wie schon weiter oben erwähnt) zum sicheren Zugriff auf Daten um Verschlüsselungsmethoden erweitert werden, die es erlauben Statements in feingranularer Weise zu verschlüsseln, was ein

weitgehend offenes Forschungsfeld darstellt, siehe etwa (Fernández et. al., (2017)) und die darin enthaltenen Literaturhinweise für Startpunkte.

# 6 Zusammenfassung und wirtschaftspolitische Empfehlungen

Die wichtigste Ressource der Welt ist nicht mehr Rohöl, sondern Daten - so der Titel des Economist vom 6. Mai 2017. Dieser Aufmacher bringt die derzeitige Bewertung von Big Data gut zum Ausdruck, wenn auch der Umfang von Daten stärk wächst und wiederverwendet werden können während Rohöl eine definitiv begrenzte Ressource ist. Big Data - ein Begriff für den keine allgemein akzeptierte Definition vorliegt - wird hier pragmatisch als eine große Datenmenge gesehen, deren Analyse den Einsatz von Tools erfordert, die über die klassischen Anwendungsprogramme (z.B. Excel) hinausgeht. Die Erfassung, Speicherung, Analyse, Wartung, Suche, Verteilung, Übertragung, Visualisierung, Abfrage, Update und der Datenschutz sind dabei aufgrund der Größe des Datenbestands Herausforderungen. Ziel ist es Strukturen in den Daten zu finden die beobachtetes Verhalten erklären oder Verhalten vorhersagen können[26]. Dabei geht es weniger um Korrelation zwischen Daten, als um kausale Zusammenhänge, ebenso um die verwendeten Algorithmen, den darauf basierenden automatisierten Entscheidungen und deren Schwachstellen. Big Data Anwendungen sind nicht auf die Wirtschaft beschränkt, sondern im militärischen Bereich, bei der Verbrechensbekämpfung, Verkehr, Landwirtschaft, Gesundheit oder in der Wissenschaft von großer Bedeutung. Die Regeln für den Umgang mit Big Data sind nicht einheitlich.

Big Data erhält ausgesprochen viel Aufmerksamkeit, weil

- es mittlerweile technisch möglich ist, das offline und online Verhalten von Personen oder Unternehmen über Datenspuren im Netz praktisch global zu erfassen, zu speichern und zu analysieren. Über die technischen Möglichkeiten dazu verfügen einerseits Geheimdienste, andererseits einige große US-amerikanische und (potentiell auch) chinesische Unternehmen. Umfassende Überwachung ist aber auch auf regionalem Niveau schon jetzt alltäglich, wenn beispielsweise Handelsketten ihre KundInnen „offline“ mit Gesichts/Personenerkennung, WLAN- und Bluetooth-Tracking verfolgen und diese Daten mit den „online“ Personen über die Bankomat- und Kreditkarte verbinden. Darüber hinaus sammeln und handeln eine Vielzahl von Unternehmen mit personenbezogenen Daten.

- in Wirtschaft und Gesellschaft bei vielen Transaktionen sensible Daten entstehen, die eine den Datenschutzbestimmungen entsprechende Behandlung verlangen. Adäquate Technologien und Prozesse werden daher nicht nur von „Datenmultis“ verlangt, sondern von allen Personen und Unternehmen die mit personenbezogenen Daten umgehen. Angewandter Datenschutz ist daher kein Randgruppenphänomen, sondern betrifft eine breite Gruppe an Akteuren.

- die Digitalisierung aller Gesellschaftsbereiche und die sinkenden Kosten für die Datenerfassug es deutlich einfacher macht, das Verhalten von Personen zu erfassen aber auch wirtschaftliche, soziale, administrative und technische Vorgänge mit weniger Aufwand digital abgebildet werden können. Neue digitale Produkte und Dienstleistungen, das „Internet of Things“ (IoT), offline/online Kunden-Tracking im Handel, smart homes, smart cities, die öffentliche Verwaltung, die digitalisierte Medizin sowie die Digitalisierung von (traditionellen) Produktions- und Vertriebsprozessen (Stichwort: Industrie 4.0)

[26] Definition in Anlehnung an Wikipedia: https://en.wikipedia.org/wiki/Big_data

erlauben es, viele Dinge im Detail zu erfassen, zu überwachen, zu steuern etc., und produzieren dabei Daten, die geradezu danach rufen, ausgewertet zu werden.

Die landläufige Vorstellung, dass man im über die Zeit aufgebauten Datenpool ohne Einschränkungen in methodischer oder technischer Hinsicht nach verwertbaren Informationen fischen kann, beflügelt viele Big Data-Fantasien und Strategien. Anläufe, das vorhandene Material zu sichten, strukturieren, vervollständigen oder die Sammelaktivitäten auszuweiten, sind vielfach Teil von Unternehmensstrategien. Dieser Ansatz muss mit dem Inkrafttreten der Datenschutzgrundverordnung (DS-GVO 2016/679) in Europa im Mai 2018 überarbeitet werden. Die DS-GVO ist - so viel kann vorweg festgehalten werden - ein strenges, europaweit einheitliches Regime für den Umgang mit personenbezogenen Daten. Folgende Herausforderungen und Auflagen kommen damit auf die Verarbeiter von personenbezogenen Daten zu:

- Explizite und informierte Zustimmung zur Verarbeitung von Daten, wobei schon das Speichern oder Anonymisieren der Daten als Verarbeitung gesehen wird. Praktisch jede Manipulation der Daten stellt demnach eine Weiterverarbeitung dar, für welche die explizite Zustimmung der Person, deren personenbezogenen Daten betroffen sind, notwendig ist.

- Die Daten - und das ist im Zusammenhang mit Big Data interessant - können analysiert werden, wenn sie anonym weiterverarbeitet werden. Für die Anonymisierung ist - wie schon erwähnt - die Zustimmung notwendig.

- Die DS-GVO bringt auch das Recht auf „Vergessenwerden“ und damit die Löschung von Daten und Informationen mit sich, ebenso wird Transparenz bei der Datenverarbeitung und Datensparsamkeit - wonach eine grundlose Speicherung auf Vorrat nicht erlaubt ist - verlangt. Diese Forderungen - so sehr sie auf einer abstrakten Ebene nachvollziehbar und verständlich sind - bringen beachtliche Herausforderungen in juristischer/legistischer und technischer Hinsicht mit sich. Zum einen sind diese Tatbestände nicht hinreichend definiert - auch weil sich die technischen Möglichkeiten laufend ändern und sich die tatsächliche Interpretation erst aus Gerichtsentscheidungen ergibt. Ein nicht unproblematischer Umstand, wenn man bedenkt, dass Verstöße gegen die DS-GVO mit einer Strafe von 4% des Umsatzes oder maximal €20 Millionen bestraft werden können. Andererseits bestehen Zielkonflikte zwischen diesen Vorgaben. Beispielsweise können sich die Forderung nach Transparenz und das Recht auf Vergessenwerden widersprechen. Hinzu kommt, dass auch die technische Umsetzung vieler Vorgaben noch Forschung benötigt, um definitive Aussagen zu den Wirkungen zuzulassen.

Die DS-GVO etabliert damit strengere Regeln als sie beispielsweise in weiten Teilen der USA gelten. Schon in der Entstehungsphase des Gesetzestextes wurde eingewandt, dass strenger Datenschutz Innovation behindert. Diese Kritik ist auch nach der Verabschiedung nicht abgeebbt und nicht zuletzt das Kernthema dieser Arbeit. Aus der Innovationsforschung ist aber bekannt, dass Regulierungen - wie beispielsweise die DS-GVO - durchaus auch zu Innovationen beitragen können. Strenge Auflagen durch die Umweltgesetzgebung haben das Wirtschaftswachstum nicht negativ beeinflusst, aber die Nachfrage nach innovativen Umwelttechnologien erhöht. Dadurch wird ein Markt für Umwelttechnologien geschaffen und potentielle Produzenten für die nunmehr nachgefragten Umwelttechnologien hatten Anreize hier innovative Angebote zu entwickeln. Kann ähnliches auch durch die DV-GVO passieren?

Die Frage lautet also, ob strenger Datenschutz Unternehmen bei Innovationen behindert oder ob er dazu beiträgt, dass sie neue Angebote, Business Modelle und Technologien entwickeln, die sowohl die regulatorischen Vorgaben als auch die Wünsche der AbnehmerInnen erfüllen. Grundsätzlich darf davon ausgegangen werden, dass die meisten NutzerInnen durchaus Wert auf Datenschutz legen, auch wenn Sie selten große Bürden auf sich nehmen, um dieser Forderung Nachdruck zu verleihen.

Um Evidenz zur Klärung dieser Frage zu präsentieren, wurde folgende Vorgangsweise gewählt:

1. Die rechtlichen Bestimmungen werden auf ihre technischen Implikationen hin untersucht. Dabei solle herausgearbeitet werden, ob es bereits Technologien und Strategien gibt, die dabei helfen explizite Zustimmung einzuholen, das Recht auf auf Löschung umzusetzen, Daten zu anonymisieren um dann Big Data-Auswertungen ohne Einschränkungen laufen zu lassen, und die Transparenzanforderungen zu erfüllen.

2. Der Bewertung der technischen Möglichkeiten bildet den Ausgangspunkt für die Abschätzung der Wirkungen auf Innovationen. Dabei wird vor allem versucht, jene Faktoren zu identifizieren die sowohl positive als auch negative Effekte auf Innovation haben können. Diese Aufgabe zwingt zu einer sehr pragmatischen Vorgangsweise, weil die Effekte der DS-GVO natürlich erst beobachtbar sind, wenn diese ab Mai 2018 unmittelbar anwendbar ist und auch dann werden die Wirkungen erst nach mehreren Jahren sichtbar.

3. Basierend auf diesen Analyseschritten werden Empfehlungen für die Wirtschaftspolitik abgeleitet.

## 6.1 Zusammenfassung der Erkenntnisse zur DS-GVO

### 6.1.1 Explizite Einwilligung - ex ante, ex post und darüber hinaus

Das **Einholen expliziter Einwilligung** zur Verwendung oder Weitergabe von personenbezogenen Daten von betroffenen Personen ist ein zentraler Bestandteil der DS-GVO. Die Einwilligung kann elektronisch eingeholt werden, wobei der Verwendungszweck klar hervorheben ist. Die Verarbeitung personenbezogener Daten darf nicht über diesen Verwendungszweck hinausgehen. Ausnahmen existieren in Bezug auf die Verarbeitung zu im öffentlichen Interesse liegenden Archivzwecken, zu wissenschaftlichen oder historischen Forschungszwecken und zu statistischen Zwecken.

Im folgenden Kapitel werden die mit dem Einholen von explizitrer Einwilligung verbundenen technischen Herausforderungen analysiert.

Zustimmung ist auf verschiedene Arten möglich: 1. Generelle Zustimmung wie zum Beispiel derzeit üblich bei den Nutzungsbedingungen von Webseiten oder 2. Opt in, die explizite Einwilligung eine bestimmte Dienstleistung nutzen zu wollen.

Die Bestimmungen der DS-GVO bringen folgende Herausforderungen mit sich:

- Kategorisierung: Um die BenutzerIn nicht zu überfordern oder mit unnötig vielen Interaktionsschritten zu konfrontieren, geht man derzeit davon aus, dass Anfragen um Einwilligung analog zu den verschiedenen Verarbeitungsschritten gruppiert werden und per Kategorie nach Einwilligung gefragt wird.

- Anpassung und Widerruf: Die AnwenderIn muss dann in der Lage sein - entgegen dem status quo, der nur Zustimmung oder Ablehnung erlaubt - die Einwilligungen partiell und feingranular zu verwalten. Damit ist auch die Möglichkeit verbunden, schon gegebene Einwilligungen zu widerrufen.

- Verständlichkeit: Die Einwilligungserklärungen/-anfragen müssen in verständlicher Form präsentiert werden und die BenutzerIn nicht überfordern. Dies ist speziell in Big Data-Szenarien herausfordernd, weil dort komplexe Datenverarbeitungsschritte, die auf nicht-trivialen Algorithmen und Verarbeitungsschritten basieren, erklärt werden müssen. Hier ist es schwierig das Mittelmaß zwischen Überforderung und Abstraktion zu finden. Es soll aber den interessierten BenutzerInnenn dennoch möglich sein, zu verstehen, was ein Datenverarbeitungs-Algorithmus mit ihren personenbezogenen Daten macht, und wie das dem eingewilligten Verwendungszweck entspricht, bzw. auch nachvollziehen zu können, welche Anonymisierungsverfahren und Sicherheitsmaßnahmen angewandt werden, um Daten zu schützen. Vorstellbar sind hier verschiedene Abstraktions- bzw. Detailebenen.

- Nachträgliche Einwilligung: Ein weiterer Aspekt für datenverarbeitende Firmen ist die Unantastbarkeit von Daten für neue Verwendungszwecke für deren Verwendung noch keine explizite Einwilligung existiert. Daher muss man technische Möglichkeiten schaffen, dass Einwilligungen unkompliziert (für beide, den Datenverarbeiter und die BenutzerInnen) eingeholt werden können.

Diese Einwilligungserklärungen sind natürlich unterschiedlich, je nachdem welche Verarbeitungsschritte gesetzt und welche Daten dabei verwendet werden. Die Konzeption einer DS-GVO gerechten Einwilligungserklärung ist daher nicht trivial und stellt für viele Anbieter eine veritable Herausforderung dar. Zuallererst sind hier die Interessenvertretungen gefragt, ihre Mitglieder zu unterstützen.

Die präzise Darstellung der Verarbeitungsschritte ist natürlich eine Herausforderung, wenn Innovationen entwickelt werden, weil es dort ausgesprochen schwierig ist, ex ante anzugeben, welche Schritte gesetzt werden. Von Seiten der Datenverarbeiter ist es wünschenswert, dass die Einwilligung zur Datenverarbeitung soweit wie möglich auch die Entwicklung neuer innovativer Dienste und Business Intelligence-Lösungen ermöglicht, ohne jedes Mal den Benutzer mit neuen detaillierten Einwilligungsanfragen konfrontieren zu müssen. Aber dazu weiter unten mehr. Von Seiten der NutzerInnen wäre es wünschenswert, wenn der Gesetzgeber hinsichtlich der Einschränkungen der Dienstleistungen vorgibt, wenn BenutzerInnen die Einwilligung nicht oder nur teilweise erteilen bzw. zurückziehen, um zu garantieren, dass ein mehr an Datenschutz nicht zu einem unfairen, diskriminierend eingeschränkten Zugang zu Dienstleistungen führt.

Die DS-GVO macht klar, dass informierte und spezifische Einwilligungserklärungen eingeholt werden müssen. In rezenten Forschungsarbeiten sind Zweifel aufgekommen, ob dies tatsächlich möglich ist. Zum einen zeigt sich, dass kurze Pausen bei der Präsentation der Datenschutzerklärungen/-bedingungen, die zur Kommunikation irrelevanter Informationen genutzt werden, genügen, um die Aufmerksamkeit für bzw. das Verständnis von Datenschutzerklärungen/-bedingungen zu vermindern oder gar zu eliminieren. Auch tendieren NutzerInnen dazu, solchen Bedingungen zuzustimmen ohne sich im Klaren zu sein, welche ihrer Daten gesammelt werden, sodass ihre Einwilligung nicht als informierte Zustimmung zu werten ist. Andererseits wurde beobachtet, dass ein Mehr an Kontrolle zu fahrlässiger Handhabung und höherer Risikobereitschaft führt. Diese Einsichten legen nahe, dass es hier durchaus noch

Forschungsbedarf gibt, wie man tatsächlich zu einer informierten Einwilligung kommen kann. Gleichzeitig heißt das aber auch, dass auch bei besten Absichten und tadelloser Umsetzung das Ziel einer bewusste Einwilligung zur Verwendung von personenbezogene Daten nicht einfach zu erreichen ist.

Anstatt monolithischer, statischer Einwilligungserklärungen besteht eine Notwendigkeit, Technologien zu entwickeln, die die dynamische Anpassung von Zustimmung erlauben, mit speziellem Fokus auf rechtliche und auch ethische Aspekte sowie einfache Benutzbarkeit und Verständlichkeit. Solche Technologien und Mechanismen sollten es einerseits Benutzern ermöglichen Daten zu korrigieren, die Nutzung ihrer Daten nachzuvollziehen und anzupassen.[27]

Begrüßenswert wäre hier die Schaffung von „Best Practices“ bzw. eines europäischen Standards für maschinenlesbare Einwilligungen bzw. Verarbeitungsschritte, Nutzungsszenarien und Kategorien. Dies würde auch zu einem weiten Feld an Drittlösungen (Apps, Bots, ...) führen, mit deren Hilfe BürgerInnen Einwilligungen automatisieren bzw. vereinfachen könnten.

Dahingehend sind etliche der weiter oben beschriebenen maschinenlesbaren Policy-Sprachen bereits in RDF (Resource Description Framework), ein Standard format des World Wide Web Consortium, zum Austausch von maschinenlesbaren Metadaten) modelliert, die zur Beschreibung und Verifikation von Nutzungsstrategien, rechtlichen Regularien und Geschäftspraktiken verwendet werden können und mit Datenprovenienz-Informationen und Transparenzinformationen und -ereignissen verknüpft werden können zur automatischen Verifikation von End-User Abkommen. Dennoch existieren noch Lücken, da für all diese Aspekte *verschiedene* Standard RDF Schemata/Vokabulare existieren und für deren gemeinsame Verwendung zur transparenten Beschreibung der Aspekte der Verarbeitung von personenbezogenen Daten und damit verbundenen Policies wiederum Standards und Best Practices fehlen.[28]

Außerdem existieren im Hinblick auf die Skalierbarkeit offene Forschungsfragen zur notwendigen Ausdrucksstärke solcher Policy-Sprachen und zur Komplexität und Skalierbarkeit entsprechender Verifikationsmethoden. Vokabulare zur Beschreibung von Strategien wie ODRL, die derzeit von der Permissions and Obligations working group des W3C standardisiert werden, leiden nach wie vor unter teilweiser semantischer Ambiguität, die einer Maschinenverarbeitung im Weg stehen, bzw. könnten sie sich in der Praxis als unvollständig oder unzureichend entpuppen, um komplexe Strategien zur Verarbeitung personenbezogener Daten zu beschreiben. Zusätzlich muss RDF zum sicheren Zugriff auf Daten um Verschlüsselungsmethoden erweitert werden, die es erlauben Statements in feingranularer Weise zu verschlüsseln, was ein weitgehend offenes Forschungsfeld darstellt.

### 6.1.2 Das Recht auf Vergessenwerden

Das in der DS-GVO stipulierte Recht auf Löschung von Daten ermöglicht es in gewissem Rahmen sensible und personenbezogene Daten aus datenverarbeitenden Umgebungen und Anwendungen zu löschen. So simpel diese Vorschrift erscheinen mag, eröffnet sie ein

[27] Hier existieren mögliche Ansätze aus dem Bereich Linked Data: Kirrane et al. geben in ihrem Survey einen Überblick zu Access Control Techniken für RDF.

[28] Siehe dazu auch https://www.specialprivacy.eu/images/documents/SPECIAL_D6.3_M9_V1.0.pdf , Kapitel 4.

herausforderndes und gleichzeitig auch rechtlich unscharf definiertes Feld, weil der Terminus „Löschen“ in Datenanwendungen unterschiedlich verwendet wird. In vielen Datenanwendungen und Produkten wird dabei keine endgültige Vernichtung („physikalisches Löschen“) der Daten gemeint, sondern lediglich eine Entfernung aus der Informationsverarbeitung („logisches Löschen“), d.h. der Speicherplatz an dem die zu löschenden Daten stehen wird als frei markiert.[29]

Forensische Tools, die solcherart „gelöschte“ Dateien wiederherstellen können sind jedoch schon seit vielen Jahrzehnten verbreitet. Der Erfolg einer Wiederherstellung hängt im Wesentlichen davon ab, ob die Blöcke schon wieder genutzt wurden. Dies wiederum hängt von der seit der Löschung vergangenen Zeit und der Nutzungsintensität im Sinn der Speicherung neuer Daten auf dem Speichermedium ab.

Um diese Wiederherstellung unmöglich zu machen, wurde in der Vergangenheit ein Set an Möglichkeiten der „endgültigen“ Löschung entwickelt, von denen das Überschreiben der Speicherbereiche mit Zufallswerten die populärste ist. Für diese Form gibt es auch eine umfangreiche Sammlung an Tools. Obwohl es in diesem Bereich durchaus noch Diskussionen über die richtige Form des Überschreibens gibt (random, patterns, mehrfach), da nach einmaligem Überschreiben mit fixen Mustern in akademischen Labors noch ursprüngliche Daten wiederhergestellt werden konnten, gilt diese Methode in der praktischen Anwendung als hinreichend sicher um von einer endgültigen Löschung der Dateien ausgehen zu dürfen.

Allerdings hat sich in extrem kritischen Bereichen durchaus die physische Zerstörung von Datenträgern eingebürgert. Dies ist speziell dann sinnvoll, wenn eine lange vorgehaltene große Datenmenge als Gesamtheit endgültig vernichtet werden muss. Dieser Zugang ist allerdings für die Umsetzung eines Rechts auf Vergessenwerden, bei dem es lediglich um die Löschung einzelner oder einer kleinen Menge an sensiblen Informationen geht, wirtschaftlich nicht vertretbar und in hochverfügbaren Systemen auch technisch nicht realisierbar.

Basierend auf den in diesem Projekt durchgeführten Analysen haben sich die untenstehenden Forschungsfragen als wesentlich herauskristallisiert. Dabei handelt es sich nicht nur um rein technische Fragestellungen, sondern vor allem um Fragen, die einen integrierten Forschungsansatz zwischen technischen und rechtswissenschaftlichen Experten erfordern:

- Welche Form des Löschens ist ausreichend und welche forensischen Methoden existieren? Dies umfasst auch die Entwicklung neuer forensischer Methoden, die einfach umzusetzen sind und speziell in sehr komplexen Systemen vorhandene, nicht gelöschte, Metainformationen zur Datenrekonstruktion ausnutzen. Dies ist extrem relevant, um den durch die DSGVO implizierten Schutz der Daten durch Löschen auch in der Realität umzusetzen. Dabei ist die Frage nicht nur auf „physikalisches“ oder „logisches“ Löschen beschränkt, sondern umfasst auch den Umgang mit Backups, Sicherheitsmechanismen, internen (Security-)Logs, sowie anderen Methoden fortschrittlichen Datenmanagements.

- Zu klären ist auch der Zielkonflikt in Hinblick auf die Transparenz der Datenverarbeitung. Da es auch notwendig sein kann Löschungen rückgängig zu machen, muss der gelöschte Inhalt in entsprechenden Mechanismen vorgehalten werden. Bestimmte Regulierungen verlangen, dass Daten nicht gelöscht werden, damit man die Entscheidungen zu einem bestimmten Zeitpunkt nachvollziehen kann. Die gelöschten Zellen werden in der

[29] Das Problem ist, dass Daten zwar nicht mehr logisch vorhanden sind - da nicht mehr auf sie zugegriffen werden kann - aber immer noch auf dem Speichermedium vorhanden sind und daher von einem Angreifer abgegriffen werden können. Hier bietet sich grundsätzlich Verschlüsselung als Option an, die allerdings bei Altsystemen (legacy systems) nicht zur Anwendung gebracht werden kann.

Datenbank in einem eigenen Index verwaltet, der sogenannten Garbage Collection, und somit nicht nur bezüglich ihres Inhalts, sondern auch der Lösch-Timeline analysiert. Wie mit diesen Zielkonflikten umzugehen ist, sollte geklärt werden.

Grundsätzlich ist jedoch das Problem des Löschens nicht alleine auf Datenträger beschränkt und kann in vielen Fällen auch gar nicht rein auf die Zerstörung von Datenträger bezogen betrachtet werden. Ein wesentlicher Aspekt ist die Speicherung in der Cloud oder externen Rechenzentren, bei denen der physikalische Speicher nicht unter der Kontrolle des Datenbesitzers liegt. Hier ist oftmals auch eine physische Löschung durch Überschreiben nur sehr schwierig zu realisieren, da die zugrundeliegende Architektur nicht bekannt und oftmals lediglich emuliert ist und entsprechende Lösch-Software nicht sinnvoll angewandt werden kann. Ähnliches trifft auch für das Löschen aus Backups zu, Bandspeicher, oder andere Massenspeicher, sind nicht darauf ausgelegt, dass einzelne Datensätze gelöscht werden und bieten diese Möglichkeit auch nicht immer an. Zusätzlich ist dies auch ein rechtliches und organisatorisches Problem, da Backups oftmals nicht angegriffen werden dürfen und speziellen Regularien unterliegen, die es einzuhalten gilt, um diversen Security-Zertifizierungen zu genügen. Hier sind auch auf organisatorischer Ebene neue Wege und Methoden zu finden, wie mit solchen Szenarien umzugehen ist,

### 6.1.3 Anonymisierung - Big Data ohne Einschränkungen?

Anonymisierung von sensiblen Daten gewinnt durch die DS-GVO an Bedeutung, weil sie eine Alternative zur Einholung von expliziter Zustimmung zur Verwendung von personenbezogenen Daten darstellt. Allerdings muss festgehalten werden, dass auch die Anonymisierung von Daten eine Verarbeitung ist und daher expliziter Zustimmung bedarf.

Wesentlich für die Sicherstellung der Anonymität ist dabei eine genaue Analyse der in den Daten enthaltenen Informationen in Hinblick auf die Möglichkeit, aus scheinbar unpersönlichen Informationen Personen eindeutig identifizieren zu können. Dabei werden die Daten grundsätzlich in drei Typen eingeteilt: Identifizierende Daten, quasi-identifizierende Daten - d.s. Daten die für sich gestellt unproblematisch sind, in Kombination jedoch die Identifizierung ermöglichen-, und die Nutzungsdaten. Bei der Anonymisierung geht es hauptsächlich um die ersten zwei Gruppen.

Es gibt eine Reihe von Strategien und Methoden, die die Anonymisierung von Daten erlauben. Das Spektrum reicht von synthetischen Daten, Katastern u.a., k-anonymisierten Daten und davon abgeleitete Verfahren bis zu differential privacy.

Ohne an dieser Stelle auf die verschiedenen Methoden einzugehen kann festgehalten werden, dass es bei einem Teil der Daten zu einen Zielkonflikt zwischen starker Anonymisierung und dem Informationsgehalt der Daten kommt. Je stärker die Anonymisierung, desto geringer ist der Informationsgehalt der Daten und damit deren Nutzen für analytische Zwecke. Allerdings sind die Nutzungsdaten – d.h. ohne personenbezogene Daten – für viele Big Data Anwendungen völlig ausreichend.

Eines der Hauptprobleme beim praktischen Einsatz von Anonymisierungsverfahren ist das Fehlen – auch bei der DS-GVO - exakt definierter rechtlicher Anforderungen an die Stärke der Anonymisierung (z.B. entspricht im Fall von k-anonymity der Faktor k der Mindestgröße der Äquivalenzklassen). Hinzu kommt, dass Datenmanipulationen, die heute Anonymität sichern, mit

fortschreitender technologischer Entwicklung “geknackt” werden können und dann so nicht mehr zulässig sein würden. Anonymisierung ist also ein “moving target”.

- Zusätzlich sind auch die folgenden Fragestellungen zu beachten, die beim praktischen Einsatz von Anonymisierungsverfahren auftreten und in der Theorie derzeit noch nicht ausreichend betrachtet wurden:Die Wahl der konkreten Sicherheitsparameter zur Anonymisierung, speziell des Sicherheitsfaktors „k“ im Rahmen von k-anonymity oder verwandten Verfahren. Das gleiche gilt auch für den Einsatz von Differential Privacy, hier ist die Wahl des Faktors Epsilon zu klären. Im Fall von Datenperturbation, d.h. der Verschneidung von Echtdatensätzen mit synthetischen Daten, ist zu klären, ab welchem Verhältnis zwischen Echtdaten und Perturbationsdaten die Privacy der beteiligten Personen gewahrt bleibt.
- Speziell im Fall von Sensordaten kann die Einteilung, ob es sich bei den Daten um sensible Informationen handelt, nicht trivial sein. Hier wird, unter Umständen branchenspezifisch, zu klären sein, wodurch sich Quasi Identifier auszeichnen und generelle Kriterien festzulegen, wie diese zu erkennen und mit ihnen umzugehen ist.
- Entstehen im Rahmen interner Datenverarbeitung sensitive Datenströme durch die Verschneidung von (u.U. teilweise sensiblen) Daten, so wäre zu klären, ab wann eine Anonymisierung durchgeführt werden muss, bzw. ob der Akt der Verschneidung noch unanonymisiert durchgeführt werden darf. Dies ist speziell wichtig, da eine Verschneidung anonymisierter Daten oftmals nicht möglich ist.

Geringer Informationsgehalt bedeutet, dass die Daten für Big Data Analysen und für Innovationsprozesse einen deutlich geringeren Wert haben. Dabei geht es nicht um die nicht vorhandene Zuordenbarkeit zu Datensubjekten, sondern um den geringen Informationsgehalt der Daten und den daraus resultierenden geringen analytischen Wert.

Diese Faktoren bewirken, dass der Einsatz von Anonymisierungstechnologien, mit relativ vielen Unwägbarkeiten verbunden ist. Hinzu kommt, dass die verschiedenen Ansätze ein relativ hohes Niveau an Expertise verlangen, das oftmals in Klein- und Mittelbetrieben nicht vorhanden ist und so ein weiterer limitierender Faktor ist.

### 6.1.4 Transparenz - Was ist exakt wann passiert?

Die Forderung nach einer transparenten Verarbeitung der Daten entspringt direkt der DS-GVO und ermöglicht damit dem Besitzer der Daten eine sehr weitreichende Kontrolle über die Verwendung der Informationen. Zusätzlich kann abgeleitet werden, ob wirklich nur die angegebenen Daten und Informationen für eine datengetriebene Anwendung verwendet wurden.

Allerdings kann diese Forderung auch mit dem Recht auf Vergessenwerden kollidieren, speziell wenn die Forderung nach Transparenz aufgrund anderweitiger Regularien begründet wird. Regularien wie SOX (Sarabanes Oxley Act) und Basel II stellen die Integrität der in einer datengetriebenen Verarbeitung verwendeten Daten sicher, d.h. sie garantieren, dass die verwendeten Daten zu keiner Zeit manipuliert wurden. Dies gilt speziell auch für externe Anreicherungsinformationen, sodass hiermit auch eine Wiederverarbeitung (reprocessing) ermöglicht wird, d.h. es ist möglich Daten so zu verarbeiten, wie das an einem bestimmten Zeitpunkt mit den damals vorliegenden Informationen gemacht worden wäre. Dies ist speziell wichtig in Billing-Workflows und generiert eine gewisse Beweisbarkeit gegenüber Forderungen und Anfechtungen, ist daher speziell im Bereich der Finanztransaktionen von hoher Bedeutung.

Wesentlich bei der Durchsetzung dieses Aspekts der DS-GVO ist besonders die Frage, welches Recht und welche Pflicht als höherwertiger anzusehen sind: Das Recht auf Vergessenwerden, oder die lückenlose Nachvollziehbarkeit, bzw. sogar eine etwaige Forderung des Wiederverarbeitung. Hierbei wird es nach unserem Dafürhalten keine allgemeingültige Entscheidung geben, sondern eine, die auf den jeweiligen Use Case und die Art der Verarbeitung abstellt.

Transparenz im Zusammenhang mit der Verarbeitung von personenbezogenen Daten kann auch insofern eine Hürde darstellen, da keine Standard-Schemata für personenbezogene Daten bzw. noch keine allgemein anerkannten „Best Practices" für entsprechende Granularität der Transparenz-aufzeichnungen existieren. Im Bereich der Forschung aus dem Bereich des Semantischen Web existieren einige Vorschläge zu Ontologien (konzeptuelle Schemata, die mittels RDF Daten instanziiert werden können) zur Beschreibung von personenbezogenen Daten und deren Provenienz (vgl. z.B. Bartolini et al. (2015)). Jedoch handelt es sich bei dieser und ähnlichen Arbeiten mehr um akademische Schemata denn um Standards die unmittelbar zum Einsatz kommen könnten. Es kann angenommen werden, dass die Entwicklung und Einführung solcher Standards zur leichteren Verarbeitbarkeit und Überprüfung von Transparenzaufzeichnungen beitragen würde.

Eine weitere technische Lösung stellt eine sogenannte Transparenzschicht dar, die mit bestimmten Eigenschaften ausgestattet ist (Vollständigkeit, Vertraulichkeit, Korrektheit, Unveränderbarkeit, Integrität, Interoperabilität, Unleugbarkeit, Richtigstellung und Löschung, Nachverfolgbarkeit/Nachvollziehbarkeit) und robuste Services (hohe Verfügbarkeit und Performanz, Skalierbarkeit und effiziente Speicherung) garantiert.

Dabei wird immer eine lokale Transparenzschicht und eine globale Transparenzschicht einer von Datenverarbeiter und Datensubjekt als sicher eingestuften Drittorganisation, oder eine global verwaltete Transparenzschicht gespeichert in einer peer-to-peer Architektur, benötigt.

Eine möglichte Architektur für eine Transparenzschicht stellt die in letzter Zeit (v.a. durch crypto-currencies) populäre Blockchain-Technologie dar, um Zugang zu personenbezogenen Daten zu managen und zu loggen. Die Blockchain-Technologie basiert per se auf peer-to-peer Netzwerken und Verschlüsselung. Eine genaue Analyse der nicht-funktionalen Aspekte von P2P Schichten (Ledgers) oder Blockchains als Basis für eine Transparenzschicht ist hier jedoch dringend nötig: Es sei erwähnt, dass speziell in Blockchains verbreitete Voting-Techniken im P2P Bereich, Manipulationen ermöglichen, wenn eine Organisation mehr als die Hälfte aller peers kontrolliert. Dies ist speziell bei privaten Blockchains bei geringer Anzahl von peers zu beachten. Desweiteren ist nachdem in einer Blockchain per se nichts gelöscht werden kann, zu klären inwiefern diese Technologie - etwa unter Verwendung von kryptographischer Löschung (also Zerstörung der Schlüssel) – gleichzeitig Transparenz und das Recht auf Löschung sichergestellt werden kann. Diese Fragen benötigen dringend noch Antworten aus Wissenschaft und Forschung und sollten mit entsprechenden Förderungen unterstützt werden.

### 6.1.5 DS-GVO gerechte Entwicklung von Big Data Anwendungen

Aufgrund der rechtlichen Anforderungen ist somit eine „naive" Entwicklung von Big Data Anwendungen im Lichte der DS-GVO nicht möglich: Datensammlungen und Data Mining Analysen ohne Einwilligung sind nicht möglich, detto die Demonstration einer Big Data basierten Anwendung anhand der eigenen Daten mit Widerrufsmöglichkeit (Opt-out).

Auf Basis dieser Erkenntnisse bietet sich allerdings folgende DS-GVO kompatible Vorgangsweise zur Entwicklung einer Big Data Anwendung an:

Bereits bei der Entwicklung der datengenerierenden Systeme wird die Einwilligung zur Verwendung der generierten Daten für spätere Analysen unter Angabe des Verarbeitungszwecks (z.B. personalisierte Werbung) eingeholt. Hierzu gibt es zwei Möglichkeiten: Entweder die Betreiber des datenverarbeitenden Systems erreichen eine Einwilligung der Datensubjekte in den datenverarbeitenden Prozess, der genau spezifiziert werden muss (Consent), oder sie erhalten Consent zur Anonymisierung und der anonymisierten Weiterverarbeitung der Daten. In letzterem Fall muss dann keinerlei genauer Verwendungszweck im Analyseschritt mehr angegeben werden und die anonymisierten Daten dürfen auch für neue, zum Zeitpunkt der Spezifikation unbekannte, Analysen verwendet werden.

Der Vorteil der ersten Variante ist, dass auf Basis der anonymisierten Daten beliebige Algorithmen angewandt werden können. Der Nachteil ist der im technischen Teil der Studie aufgezeigte Informationsverlust. Aus rechtlicher Sicht wäre es wünschenswert, wenn in der Einwilligung die „Anonymisierungsmethode lediglich die Angabe „Anonymisierung“ ausreichend bestimmt wäre, zumal das konkrete zur Anonymisierung verwendete Verfahren an den State of the Art anzupassen ist.

Da das konkret einzusetzende Data Mining Verfahren a-priori nicht bekannt ist, ist davon auszugehen, dass die Einwilligung öfter adaptiert werden muss, falls die Daten nicht anonymisiert werden. Auch hier ist die im Rahmen der Anwendung der DS-GVO zu klärenden ausreichenden Granularität der Einwilligung wichtig ist. Jedenfalls ist auch die Verständlichkeit für den Anwender zu berücksichtigen, zumal sowohl Anonymisierungsalgorithmen als auch Data Mining Algorithmen komplex sind.

Im Sinn der Datensparsamkeit empfiehlt sich in jedem Fall eine Trennung der operativ verarbeiteten Daten, die nur hierfür gespeichert werden, und der Inputs für die Big Data Analysen.

Bei der Entwicklung der darauf basierenden Anwendungen empfiehlt es sich, eine Gruppe von Testern zu requirieren, die eine auf die Entwicklung abgestimmte Einwilligungserklärung abgeben, falls die Einwilligung zur Datenanalyse nicht ausreicht. In diesem Fall sind dann beim Echteinsatz der neuen Applikation sind dann von allen NutzerInnen entsprechend abgestimmte Einwilligungserklärungen einzuholen.

Im Vergleich zur „naiven“ Big Data Entwicklung stellt diese Vorgangsweise sicher, dass die Nutzer immer über die Art und Weise wie ihre Daten verwendet werden informiert sind. Nachteile sind die Einschränkung der Analysemöglichkeiten, die Unmöglichkeit, einer AnwenderIn den Nutzen der neuen Anwendung auf Basis der eigenen Daten zu demonstrieren und die nicht mögliche Nutzung der „Power of Defaults“ durch ein nachträgliches Opt Out.

Diesen Nachteilen steht allerdings der Vorteil eines größeren Vertrauens in den Datenschutz des anbietenden Unternehmens gegenüber, der zu einer höheren Bereitschaft führt, der Datenverwendung zuzustimmen. Dies ist insbesondere im Zusammenhang mit dem Recht auf Löschung und den Transparenzregeln zu sehen, die eine jederzeitige spätere Löschung der eigenen Daten ermöglichen. Diese Möglichkeiten sprechen auch für eine großzügigere Auslegung der Bestimmtheitsanforderungen bei der Einwilligung zur Anonymisierung bzw. Datenanalyse.

## 6.2 Innovation

In einer zunehmend digitalen Gesellschaft und Wirtschaft sind der Zugang zu Daten und die damit erlaubten Handlungen ein wesentlicher Faktor, um Einsichten über ablaufende Prozesse zu gewinnen. Je besser diese analysiert, abgebildet und letztendlich prognostiziert werden können, desto größer ist auch der Wert der Daten für die Gestaltung von Interventionen in Wirtschaft, Politik, Verwaltung, Verbrechensbekämpfung etc.

Die unterschiedlichen Entwicklungsoptionen für wirtschaftliche Aktivitäten und Produkt- und Prozessinnovationen, die sich aus abweichenden datenschutzrechtlichen Bestimmungen ergeben, sind eine zentrale Fragestellung dieser Studie. Es gibt offensichtlich die Annahme, dass Datenschutzbestimmungen Einfluss auf die Entwicklung von Branchen, Unternehmen und die öffentliche Hand (eGovernment) haben und Innovationsprozesse beeinflussen. Goldfarb und Tucker (2011) sehen in den Bereichen online Werbung, eHealth und unternehmensinternen Diensten die größten Effekte.

Keine spezifische Evidenz wurde gefunden, dass strenge Datenschutzbestimmungen positive Wirkungen auf Innovation hätten. Das liegt natürlich auch daran, dass die neue DS-GVO noch nicht in Kraft ist. Dennoch sehen viele Beobachter eine Positionierung Europas als sicherer Hafen für personenbezogenen Daten als Entwicklungschance für die digitale europäische Wirtschaft. Im günstigsten Fall erzwingt das striktere Datenschutzregime in Europa Geschäftspraktiken, die die Akzeptanz europäischer Produkte und Leistungen erhöhen und damit Wettbewerbsvorteile kreieren. Wenn man vermeintliche Nachteile als Herausforderung betrachtet, entsteht Raum für den kreativen Umgang mit den Beschränkungen und für neue Lösungen.

Es geht vor allem um technische Lösungen, die sowohl die Privatsphäre gewährleisten und dennoch die Analyse der Daten erlauben und damit Big Data-Anwendungen ermöglichen. Vorrangig ist dabei die Entwicklung von Technologien, die nur wenige Daten benötigen, um eine gewünschte Funktionalität zur Verfügung zu stellen, die Möglichkeit Daten selektiv zu löschen und die Anonymisierung von Daten (siehe dazu weiter oben).

Vorweg stellt sich die Frage, wie die DS--GVO Innovationsprozesse beeinflussen kann. Im Wesentlichen wurden die folgenden 5 Wirkungsketten gefunden, über die Datenschutz Innovationsaktivitäten beeinflussen kann[30]:

1**. Produkt- und Dienstleistungsinnovation**: Innovationsprozesse sind Suchprozesse in denen mit dem Umfeld interagiert wird. Sie sind weitgehend offene Prozesse. Die Digitalisierung hat es tendenziell leichter gemacht relevante Expertise hereinzuholen – Stichwort Open Innovation. Innovationsprozesse sind tendenziell datengetriebener als in der Vergangenheit. Ansätze wie Lean Startup propagieren die Bildung von Hypothesen und deren datenbasierte Validierung durch potentielle Kunden über den ganzen Innovationsprozess hinweg. Auch war es schon bisher üblich Marktforschung und interne Datenquellen zu nutzen, um Einsicht in das Nachfrageverhalten und Kundenwünsche zu erhalten.

Grundsätzlich könnte hier die neue Datenschutzgrundverordnung Einschränkungen bringen, weil ein Teil der Datensubjekte keine Zustimmung zur Verwendung der Daten gibt bzw. weil für die Nutzung historischer Daten eine neuerliche Einwilligung notwendig ist – dazu mehr weiter unten.

[30] Nicht behandelt wird hier eHealth – der Gesundheitsbereicht – bei welchem strenger Datenschutz eine Voraussetzung für die digitale Nutzung der Daten ist. Siehe dazu beispielsweise Goldfarb – Tucker (2013).

In der Praxis kann man dabei zwei Fälle unterscheiden die unterschiedlich zu bewerten sind: Startups und etablierte Unternehmen.

Startups – bei denen das Projekt gleichzeitig das Unternehmen ist – sind bei Ihren Entwicklungsaktivitäten auf Daten angewiesen bzw. propagiert die Lean Startup-Methode eine datengetriebene Vorgangsweise[31]. Der Umfang und die Erhebungsmethoden sind allerdings selten als Big Data zu bezeichnen. Zumeist handelt es sich um persönliche Interviews, Umfragen, Nutzungsdaten von Webseiten etc., worin durchaus sensible Informationen enthalten sein können, aber es in der Regel auch möglich ist, die Zustimmung der Datensubjekte für die Verwendung einzuholen. Bei Startups ist also nicht grundsätzlich davon auszugehen, dass die DS-GVO Innovationsprozesse nachhaltig behindert.

Etablierte Unternehmen haben funktionierende Business Modelle und damit oft einen beachtlichen Bestand an historischen Daten. Innovationsprozesse sind auch hier zunehmend von lean startup, design thinking[32] und behavioural economics beeinflusst und damit auch sehr stark von der Interaktion mit den Endkunden getrieben. Hier ist die Problematik ähnlich wie bei startups.

Allerdings gibt es für etablierte Unternehmen eine Reihe von Optionen, um mit den strengeren Vorgaben der DS-GVO umzugehen. Zum einen können Tests mit internen Usern (MitarbeiterInnen) gemacht werden oder eine Gruppe von Testusern in den Entwicklungsprozess eingebunden werden. Es ist daher schwer zu argumentieren, dass Datenschutz die Entwicklung von Produkt- und Prozessinnovationen grundsätzlich behindert. Diese simplistische Aussage findet sich auch in keinem der hier analysierten wissenschaftlichen Artikel. Es ist daher naheliegend, dass die DS-GVO Anpassungen in Innovationsprozessen verlangt, diese aber keineswegs fundamentale Einschränkungen für Innovationsaktivitäten bringen.

Gegen diese Sichtweise spricht, dass Innovationsprozesse selten wohlgeordnet ablaufen. Beispielsweise kann das in der DS-GVO festgelegte Prinzip der Datensparsamkeit funktionieren, wenn man im Innovationsprozess exakt weiß, was man wann machen will. Mit dieser hohen Voraussicht kann auch das Verhältnis mit den Datenlieferanten, deren personenbezogenen Daten verarbeitet werden sollen, geklärt werden. In der Realität ist diese Situation - wenn die Innovation riskant ist - so gut wie auszuschließen. Innovation ist als Prozess “messy”, geprägt von ständigem vor und zurück auf der Suche nach Kombinationen die auf dem Markt funktionieren (siehe dazu Abbildung 19), d.h. ein Bedürfnis in einem klar definierten Marktsegment befriedigen und dabei so viel Umsatz generieren, dass die Entwicklungskosten und die zukünftigen Betriebskosten hereingespielt werden können. Dazu braucht es zumeist mehrere Anläufe und - in nicht wenigen Fällen - auch einen sogenannten Pivot, d.h. eine grundlegende Änderung der Entwicklungsrichtung, andere Produkt- bzw. Dienstleistungseigenschaften bzw. ein anderes Marktsegment. Gerade wenn es um diese grundlegenden Veränderungen geht, braucht es eine möglichst gute Datenbasis, um die Richtung zu fixieren. Innovatoren sind daher bemüht, möglichst viele Daten zu sammeln, damit die wesentlichen Faktoren für den Erfolg eines Produkts erkennbar werden. In diesem Sinn - aber das ist natürlich keine juristisch fundierte Aussage - kann ein “Datenüberschuss” die effizienteste

[31] Natürlich gibt es auch Startups die Big Data – d.h. Produkte und Dienstleistungen, die auf der Analyse von großen Datenbeständen aufbauen – im Kern ihres Business Modells haben. Diese sind jedoch nur ein relativ kleiner Teil aller Startups.

[32] Design thinking versucht nach Wikipedia (https://de.wikipedia.org/wiki/Design_Thinking) „...Lösungen zu finden, die aus Anwendersicht (Nutzersicht) überzeugend sind.“ Damit gibt es die gleiche Zielrichtung und auch in der Vorgangsweise eine Reihe von Überschneidungen mit Lean Startup.

Art und Weise sein, um eine Innovation zu entwickeln. Im Kern heißt das, dass eine enge, ex post Auslegung des Prinzips der Datensparsamkeit bei Innovationsprozessen schwierig einzufordern ist, weil man ex ante nicht weiß, welche Daten man für den Innovationsprozess braucht und daher tendenziell eine eher breite Suchstrategie wählt. Ähnliches gilt natürlich auch für Forschungs- und Entwicklungsprozesse im akademischen Bereich. Eine sehr restriktive Auslegung des Datensparsamkeitsprinzips ist bei Innovationsprozessen, Startups und im wissenschaftlichen Bereich nicht hilfreich.

*Abbildung 2121: Innovations- und Designprozesse*

Quelle: Newman

Natürlich kann man obiges Problem auch durch vorherige Zustimmung lösen, wenn diese entsprechend breit formuliert wird. Hier ist derzeit unklar, wie lange man dabei auf rechtlich festem Boden steht.

Alternativ könnte man über die Anonymisierung der erhobenen Daten Restriktionen bei der Datenverwendung vermeiden. Hier gibt es erste Forschungsergebnisse zu den Auswirkungen des Rechts auf Vergessen bzw. der Anonymisierung von Daten auf die Effizienz von Machine Learning-Ansätzen (Bernd et al. (2016)) – ein Gebiet das als hochinnovativ zu bezeichnen ist. Hier geht es vor allem um die Effekte dieser Bestimmungen auf die der Datenauswertung zugrundeliegenden Algorithmen. Dabei muss beachtet werden, dass speziell bei selbstlernenden, sog. „intelligenten“ Systemen, die Daten nicht nur verarbeitet werden, sondern auch in einem wesentlichen Ausmaß das System konstituieren, d.h. bereits verarbeitete und klassifizierte Daten dienen als Grundlage zur weiteren Analyse, als sogenannte „Wissensbasis“. Die Verfälschungseffekte, die durch eine Modifikation an der Wissensbasis auftreten, können daher wesentlich für die weitere Verarbeitung sein.

Anonymisierung hat in diesem Beispiel einen weitaus größeren Effekt auf die Nutzbarkeit der Daten, als die selektive Löschung wichtiger Merkmale. Allerdings muss in Hinblick auf die Einschränkungen und Rahmenbedingungen der Ergebnisse zur Löschung darauf hingewiesen werden, dass auch in diesem Bereich noch weitaus mehr Forschung betrieben werden muss, speziell in Hinblick auf die zu wählende Anonymisierungsmethode. Dennoch bietet dieses Ergebnis einen ersten Anhaltspunkt, in welche Richtung noch speziell Forschungsaufwand benötigt wird, um Techniken des Privacy Aware Machine Learnings (PAML) zu entwickeln, die mit herkömmlichen Techniken mithalten, oder diesen qualitativ zumindest nahekommen können. Gleichzeitig wird angedeutet, dass bei machine learning die Vollständigkeit und der Informationsgehalt der Datenbasis durchaus zu einer Effizienzminderung der Systeme führen.

2. **Werbe- und datenfinanzierte Business Modelle**: Die Zahl der Business Modelle für online Aktivitäten – diese sind hauptsächlich von der neuen DS-GVO betroffen – ist begrenzt (siehe beispielsweise Croll – Yoskovitz (2013)). Ein Teil davon verwendet Werbeeinnahmen oder verkauft Daten, um die laufenden Kosten zu decken und Innovationaktivitäten zu finanzieren. Durch die benötigte explizite Zustimmung zur Datenweitergabe, kann man davon ausgehen, dass diese seltener erteilt wird als derzeit der Fall. Ähnlich ist es auch bei der Verfolgung von NutzerInnen, um Daten zur personalisierten Schaltung von Werbung auf Webseiten zu erhalten.

Untersuchungen haben ergeben, dass schon die derzeit gültige Version der DS-GVO die Effizienz von Werbeschaltungen herabsetzt. Diese Tendenz wird durch die zukünftige DS-GVO weiter verstärkt. Es ist also davon auszugehen, dass werbefinanzierte Business Modelle weniger geeignet sein werden, um ausreichend Ressourcen zu erwirtschaften. Allerdings heißt das nicht, dass damit Werbung im Internet der Vergangenheit angehört. Weiterhin effizient kann auf spezialisierten Seiten geworben werden, weil die Interessen der BesucherInnen dort bekannt sind bzw. wenn NutzerInnen explizit die Erlaubnis zur Verwendung der Daten gegeben haben. Natürlich werden auch die Werbeanbieter Wege suchen, um sich auf die neue Situation einzustellen und ihre Business Modelle weiterzuentwickeln. Schon jetzt – trotz der strengeren europäischen Datenschutzbestimmungen – florieren die großen, überwiegend werbefinanzierten amerikanischen Plattformanbieter auch in Europa. Ein Umstand, an dem sich in absehbarer Zeit nichts ändern wird.

Abzuwarten bleibt, ob es den NutzerInnen in Zukunft leichter möglich ist, die Weitergabe ihrer Daten zu verfolgen bzw. ob es klarer sein wird, dass sie mit ihren Daten für kostenlose Produkte zahlen. In jedem Fall, kann davon ausgegangen werden, dass die DS-GVO Business Modelle mit werbe- und datenfinanzierte Innovationsaktivitäten weniger attraktiv macht.

Theoretische Modellierungen dieser Situation zeigen deutlich, dass es Anbieter unter diesen Bedingungen tendenziell Bezahlmodelle anbieten und nicht auf werbefinanzierte Produkte setzen.

3. **Datenschutz als Teil der Marketingstrategie:** Die Weitergabe von Daten bzw. die Schaltung von gezielter Werbung entsprechen schon jetzt nicht den Wünschen der NutzerInnen. Die überwiegende Mehrheit ist klar für strengen Datenschutz und Kontrollmöglichkeiten was die Weiterverbreitung von personenbezogenen Daten betrifft. Nur ein kleiner Teil ist hier bereit auch personenbezogene Daten weiterzugeben.

Das oft beobachtete Privacy-Paradoxon – dass auch User die ihre personenbezogenen Daten schützen wollen, diese bei nächstbester Gelegenheit gegen ein kostenloses Produkt eintauschen – mag als Widerspruch zum Wunsch nach Datenschutz gesehen werden. Allerdings ist zu beachten, dass es grundsätzlich sehr schwer ist, die Konsequenzen der Zustimmung zu Nutzungsbedingungen von online Diensten realistisch abzuschätzen. User versuchen daher jeweils situationsspezifisch richtige Entscheidungen zu treffen, die angesichts der spärlich vorhandenen Informationen über die Datenweitergabe kaum rational zu treffen sind. Vielfach kann man nur wählen, ob man einen Dienst nutzen will und muss damit auch den Verlust an Verfügungsgewalt über personenbezogene Daten in Kauf nehmen. Gerade bei Diensten mit starken Netzwerkeffekten wie den Sozialen Medien ist hier der Konformitätsdruck sehr stark. Gleichzeitig versuchen diese Anbieter natürlich auch zu vermitteln, dass sie sorgsam mit Daten umgehen.

Angesichts dieser Ausgangslage bringt die neue DS-GVO jedenfalls Verbesserungen, die im Interesse der KonsumentInnen sind, weil expliziter dargestellt werden muss, wie man mit den

Daten umgeht und die Einwilligung dazu eingeholt werden muss. In einem Umfeld, wo sowohl die KonsumentInnen als auch der Gesetzgeber auf mehr Datenschutz drängen, ist es naheliegend, Datenschutz zum integrierten Teil des Produktangebots zu machen und sich somit aktiv von „datenhungrigen" Anbietern zu differenzieren. Es gibt erste Anbieter die diesen Weg beschreiten.

Mit Inkrafttreten der DS-GVO im Mai 2018 ist jeder Verarbeiter von personenbezogenen Daten hier zu Anpassungen gezwungen, die es nahelegen, Datenschutzbestimmungen ernst zu nehmen und zu implementieren. Damit kann man das auch gleich zu einem Teil der Marketingstrategie machen.

Generell scheint dies eine ausbaubare Strategie für die Entwicklung von digitalen Produkten und Dienstleistungen in Europa zu sein, die auf allen Politikebene verfolgt werdeund konstituierendes Element einer horizontalen Industriepoltik sein sollte. Dies schließt natürlich nicht nur die Unternehmensseite ein, sondern sollte genauso die Leitlinie für den öffentlichen Sektor – inklusive Geheimdienste – sein. Damit könnte Europa synonym zu hohem Schutz von personenbezogenen Daten werden. Eine Positionierung die weder asiatische Anbieter noch die USA anstreben.

4. **Gesetzliche Vorgaben ermöglichen die Entwicklung von Datenschutztechnologien**: Die DS-GVO schafft grundsätzlich einen „Markt" für Datenschutztechnologien (Privacy Enhancing Technologies – PET). Diese können sowohl von den Betroffenen nachgefragt werden (VPN, TOR etc.) als auch von Unternehmen, die Datenschutz gegenüber ihren Kunden garantieren wollen. Die gesetzlichen Rahmenbedingungen sind gerade für die Entwicklung von Datenschutztechnologien wesentlich, weil dafür Nachfrage für bestimmte Produkte und Dienstleistungen geschaffen wird.

Allerdings müssen die gesetzlichen Bestimmungen hinreichend klar sein, um hier sowohl Nachfrage als auch Angebot zu stimulieren. Wenn Ziele zu viel Interpretationsspielraum lassen, steigt die Unsicherheit und nicht die Zahl der Lösungen zur Erreichung der Ziele. Konkret gibt es großen Interpretationsspielraum bei den Forderungeno zum „Löschen von Daten" oder „Transparenz" oder „Datensparsamkeit". Hier kommt es höchstwahrscheinlich erst über die Gerichte zu einer Klärung der Begriffe.

Nicht gemeint ist, dass man sich auf bestimmte Technologien festlegen solle, wenn es um den Abbau von Unsicherheit bei der Interpretation von gesetzlichen Vorgaben geht. Die Vorgaben sollten technologieneutral sein und damit lediglich klar vorgeben was erreicht werden soll, und nicht mit welcher Technologie die Ziele erreicht werden sollen.

Die DS-GVO ist aber natürlich auch ein potentieller Treiber für die Entwicklung neuer Technologien die es einerseits einfacher machen den Bestimmungen zu entsprechen bzw. für welche die DS-GVO erst einen Markt geschaffen hat. Aus derzeitiger Perspektive gibt es dabei allerdings einige Beschränkungen:

- Die postulierten Grundsätze widersprechen sich teilweise. Beispielsweise kollidiert die Forderung nach Transparenz mit dem Recht auf Vergessen. Transparenz ist nicht nur Gegenstand zahlreicher branchenspezifischer Regularien (Basel 2, SOX, HIPAA), sondern erwächst auch aus der DS-GVO selbst. Des Weiteren entsteht ein beträchtlicher Aufwand, bzw. bestehen verschiedene technische Möglichkeiten zur Realisierung der Einhaltung der in der DS-GVO geforderten Transparenzanforderungen und damit

verbundener Speicherung entsprechender Aufzeichnungen zur Verwendung personenbezogener Daten.

- Viele Bestimmungen sind nicht wirklich operationalisiert. Die Wahl der konkreten Sicherheitsparameter zur Anonymisierung, speziell des Sicherheitsfaktors „k" im Rahmen von k-anonymity oder verwandten Verfahren. Das gleiche gilt auch für den Einsatz von Differential Privacy, hier ist die Wahl des Faktors Epsilon zu klären. Im Fall von Datenperturbation, d.h. der Verschneidung von Echtdatensätzen mit synthetischen Daten, ist zu klären, ab welchem Verhältnis zwischen Echtdaten und Perturbationsdaten die Privacy der beteiligten Personen gewahrt bleibt.

- Die Klärung dieser Fragen wird Großteils von Gerichten vorgenommen werden, wobei aber nicht nur der Ausgang, sondern auch der Zeitpunkt der Klärung noch völlig offen sind.

In Summe bewirken diese Unschärfen, dass es gerade nicht zu einer Marktbildung durch die DS-GVO kommt, weil nicht abschätzbar ist, in welche Richtung sich der Markt entwickeln wird. Klar ist hingegen, dass bei vielen Fragestellungen noch relativ hoher Forschungsbedarf besteht und gleichzeitig Schritte unternommen werden sollten, die Unsicherheiten durch Umsetzungsempfehlungen zu reduzieren.

5. **Weniger Spielraum für Prozessinnovationen:** Geringere Effizienz bei unternehmensinternen Abläufen - eine wesentliche Auswirkung von mehr Datenschutz nach Goldfarb und Tucker (2012) – durch die Verhinderung von Prozessinnovationen ist eine weitere Wirkungskette. Personenbezogene Daten werden häufig nicht nur für Produktinnovationen eingesetzt werden, sondern ermöglichen es auch interne Abläufe zu neu zu gestalten und zu optimieren. Dazu gehören Werbe- und Marketingmaßnahmen ebenso wie das Rechnungswesen. Da viele dieser Leistungen auch über Drittanbieter erstellt werden, kann es hier durchaus zu Restriktionen bei der Datenweitergabe kommen (Goldfarb - Tucker, (2012)). Dies gilt vor allem dann, wenn die Datenschutzbestimmungen der Drittanbieter Bestimmungen enthalten, die die Weitergabe von Daten vorsehen.

## 6.3 Wirtschaftspolitische Empfehlungen

### 6.3.1 Ausgangslage

Mit der Einführung der DS-GVO im Mai 2018 kann festgehalten werden, dass Europäische Organisationen, die davon betroffen sind, keine wirkliche Wahlfreiheit mehr haben und Datenschutz und die daraus folgenden Beschränkungen beachten müssen, unabhängig von den Wirkungen auf Innovation. Dies gilt allerdings auch für Anbieter außerhalb Europas, sofern diese europäische KundInnen bedienen.

Die bisher analysierten Zusammenhänge zwischen Datenschutz zeigen, dass es sich um keine direkte und lineare Beziehung handelt, sondern dass auf verschiedenen Ebenen Effekte eintreten, die je nach Reaktion von Politik, Unternehmen und NutzerInnen, unterschiedliche Szenarien zulassen.

Konkret – und etwas vereinfacht – wird **kein direkter negativer Zusammenhang zwischen Innovation, Big Data und Datenschutz** festgestellt werden, weil es den Unternehmen möglich ist durch Verhaltensänderung – d.h. Anpassung an die neuen Gegebenheiten - entsprechend zu agieren. Das heißt, klar erkennen lassen, dass sie Datenschutz ernst nehmen und ein

Vertrauensverhältnis mit ihren Datenlieferanten aufbauen und auch die Verfügungsgewalt über personenbezogene Daten bei diesen belassen. Wenn dies nicht gelingt, dürften die Einwilligung zur Weiterverwendung von personenbezogene Daten wesentlich seltener gegeben werden.

Umgelegt auf die europäische Situation können daraus folgende Schlüsse gezogen werden:

Die gegenüber den USA strengeren Datenschutzbestimmungen unterstützen eine Positionierung im Bereich Bezahldienste und keine Datenmonetarisierung – d.h. den Verkauf von Daten-, weil es dazu eine explizite Zustimmung bedarf, die auch wiederrufen werden kann, und die Botschaft des besseren Datenschutzes durch den Firmenstandort zusätzliche Glaubwürdigkeit erhält.

Die Chancen von Bezahlangeboten im Privatkundenbereich mit verbesserten Datenschutz sind aufgrund des Privacy Paradoxons[33] und der geringen Marktgröße allerdings als gering einzuschätzen. Das Privacy Paradoxon wird durch die fehlende Transparenz über die Weiterverwendung der Daten verschärft. Nur wenn die Datenlieferanten wissen, wie ihre Daten weiterverwendet werden und sie die Datenweitergabe unterbinden können und dann auch Alternativen - Bezahldienste – haben, können sie die Vor- und Nachteile bewerten.

Bisherige Erfahrungen zeigen, dass das Privacy-Paradoxon gekoppelt mit den Versuchen von etablierten Plattformen sich als Unternehmen, die Wert auf Datenschutz legen, zu positionieren, dem Geschäft mit Daten nicht geschadet hat. Unternehmen wie Google oder Facebook reagieren darauf mit dem Schaffen von “walled gardens” - einem immer umfassenderen Ökosystem, das diversifizierte Dienste anbietet und damit immer weitreichenderen Zugang zu Kundendaten hat (Kelley et al, (2010)).

Chancen hätten alternative Bezahlangebote im Privatkundenbereich auf europäischer Ebene nur bei Vorliegen eines digitalen Binnenmarkts und massiver Unterstützung beim Aufbau von Netzwerkeffekten und einer verstärkten Überprüfung der Einhaltung der Bestimmungen der DS-GVO (siehe unten im Maßnahmenteil).

Europäische Regulatoren gehen zunehmend davon aus, dass es bei der Verwertung von Daten durch die dominanten amerikanischen Plattformen möglicherweise zu Marktmachtmissbrauch kommt. Ein Umstand dessen sich die Europäischen Wettbewerbsbehörden zunehmend annehmen wollen (siehe Scott – Hirst (2017)) und der tendenziell neue Chancen für alternative Anbieter eröffnen würde. Allerdings sind neue, kleinere Anbieter für viele KundInnen kaum vertrauenswürdiger als die großen etablierten Unternehmen, wenn es um Datenschutz geht. Hier bedarf es einer Reihe von Maßnahmen, um die Glaubwürdigkeit von neuen Anbietern zu stärken. (siehe unten im Maßnahmenteil)

Bessere Bedingungen für europäische Anbieter gibt es im Bereich der Unternehmen und im öffentlichen Sektor. Hier ist durch die Erhöhung der Strafrahmen im Rahmen der Datenschutz-Grundverordnung und der potentiellen Reputationsverluste bei Verstoß gegen die DS-GVO, die Zahlungsbereitschaft für Datenschutz hoch und Datenmonetarisierungsmöglichkeiten außerhalb des Anwendungskontexts oft nicht gegeben. Ebenso dürfte die Sensitivität der Konsumenten bezüglich Datenschutz im Bereich eGovernment höher als in anderen Bereichen sein.

[33] Zur Zahlungsbereitschaft für Datenschutz von Privatkunden gibt es eine Reihe von Untersuchungen, die sich mit der sogenannten. Privacy auseinandersetzen. Unter diesem Phänomen versteht man die empirische Beobachtung, dass KundInnen, die im persönlichen Gespräch angeben, sehr großen Wert auf Datenschutz zu legen, dann in einer konkreten Situation gegen eine relativ geringe Gegenleistung bereit sind, ihre Daten einem Anbieter zur Verfügung zu stellen, in dem etwa Online-Datenschutzbestimmungen ohne vorheriges Lesen zugestimmt werden (Kokolakis, (2015); Kübler, (2011)).

Verantwortliche für die Verarbeitung von personenbezogenen Daten haben in Europa für alle geplanten Verarbeitungsschritte relativ detailliert Zustimmung einzuholen oder die Daten zu anonymisieren und dann ihre Big Data-Auswertungsstrategien zu fahren. Da bei der Anonymisierung zum einen Unsicherheit über die dabei zu wählende Vorgangsweise besteht, der Umgang damit herausfordernd ist und andererseits durch die Anonymisierung ein großer Verlust des Informationsgehalts von Daten einhergehen kann, bleibt für europäische Unternehmen im Kern nur eine Option: als Standort und als Unternehmen rigoros auf die restriktiven Datenschutzbestimmungen in Europa zu setzen, damit das Vertrauen der NachfragerInnen in Europa und global zu gewinnen, und damit wettbewerbsfähig zu werden.

## 6.3.2 Maßnahmen

Wenn man der Analyse bis hierher folgt, dann braucht es eine Reihe von Maßnahmen auf Unternehmensebene, in der Technologie-, Innovations- und Industriepolitik. Natürlich ist hier auch der öffentliche Sektor gefordert.

### 6.3.2.1Maßnahmen auf Unternehmensebene

Generell braucht es mehr Awareness bei den Unternehmen. Dort dürfte nur zum Teil angekommen sein, welche Veränderungen durch die DS-GVO notwendig werden und welche Strategie aus dieser europäischen Gesetzesmaterie folgt. Demnach sollte man zwei Botschaften kommunizieren:

1. Konkrete Hinweise, welche Maßnahmen die Unternehmen treffen sollen.

2. Welche Strategien für den Umgang mit Daten erfolgsversprechend sind.

In Summe ist zu erwarten, dass europäische Unternehmen mit Big Data Ambitionen, weniger durch neue Technologien zu den gewünschten Einsichten kommen als mit einem anderen Marktauftritt, der zum Aufbau eines Vertrauensverhältnisses mit Ihren KundInnen führt und damit den restriktiveren Zugang Europas kompensiert,

Missbraucht man das aufgebaute Vertrauen, dann werden die Datensubjekte sehr viel zurückhaltender mit den expliziten Einwilligungen zur Verarbeitung von Daten umgehen. Natürlich müssen Datenschutzbehörden ausreichend Ressourcen erhalten, um Missstände verfolgen zu können bzw. auf Anzeigen zeitnah reagieren können.

Damit ergeben sich Chancen für global agierende europäische Anbieter, die spezialisierte Lösungen mit nachweislich hohem Schutzniveau für diese Kundengruppen anbieten. Ein Beispiel für eine derartige Strategie ist Fabasoft, die als erstes Unternehmen weltweit für ihre Cloud Services die mit 5 Sternen höchstmögliche Zertifizierung nach dem internationalen „EuroCloud Star Audit“ (ECSA V3.0) erhalten hat und beim Schutz personenbezogener Daten nach ISO 27018 zertifiziert ist. Dieser internationale Standard formuliert datenschutzrechtliche Anforderungen an Cloud-Anbieter. Diese müssen umfangreiche Benachrichtigungs-, Informations-, Transparenz- und Nachweispflichten erbringen, um bei KundInnen und Behörden Vertrauen hinsichtlich der Verarbeitung von personenbezogenen Daten in der Cloud zu schaffen[34].

Ein Framework, das Unternehmen bei der proaktiven Integration von Datenschutz beim Design und der Adaption von Produkten und Dienstleistungen im Bereich RFID unterstützt ist das PIA

[34] https://www.fabasoft.com/de/group/transparenz/sicherheit-datenschutz

(Privacy Impact Assessment) Framework (Oetzel, Spiekermann, (2012)). Dieses Framework baut auf Risk-Assessment-Methoden auf und bietet eine strukturierte Methode inklusive Prozess- und Dokumentvorlagen. Die Grundidee dieser Ansätze ist es, Privacy bereits in der Architektur des Angebots zu verankern und nicht nur in Policies.

Datenschutz muss in allen Aktivitäten des Unternehmens implementiert sein. Eine enge Kooperation zwischen Management, Entwicklern und Organisatoren ist daher notwendig: „Privacy by Design is designed as an engineering *and* strategic management approach that commits to selectively and sustainably minimize information systems´ privacy risks through technical *and* governance controls" (Spiekermann, (2012)). Die Unternehmen selbst sind dadurch besser gegen Hacker-Angriffe und Daten-Leaks geschützt.

Die Wirtschaftspolitik sollte darauf bedacht sein, den Umstellungsprozess auf die DS-GVO zu gestalten und handlungsrelevante Informationen zum Umgang mit den Herausforderungen anbieten. Die Zielsetzung ist dabei, die Unternehmen bei der fristgerechten und effizienten Umsetzung der DS-GVO zu unterstützen, damit man ein Vertrauensverhältnis aufbauen kann und damit die Voraussetzung für datenbasierte Innovationsstrategien legt.

Dazu gehören Anleitungen wie man die explizite Einwilligung gestalten und umsetzen kann, ebenso wie Informationen zur notwendigen Transparenz. Hier sind vor allem die Interessenvertretungen gefragt und - erst, wenn diese nicht entsprechend reagieren - die öffentliche Hand. Zusätzlich kann man Unternehmen und Startups inspirieren Kooperationen mit ausgewählten Kundensegmenten anzustreben und damit Entwicklungsarbeiten zu unterstützen. Diese Testuser sind oft bereit sowohl die mit den Tücken von noch nicht ganz ausgereiften Produkten umzugehen, als auch ihre Daten für die Weiterentwicklung zur Verfügung zu stellen.

Konkret sollte Aufmerksamkeit für die bevorstehenden Umstellungen und Herausforderungen geschaffen werden und Strategien für die Umsetzung der DS-GVO kommuniziert werden. Bei Unternehmen sind folgende Schritte Fragestellungen relevant, damit – wie schon erwähnt – Datenschutz in allen Bereich implementiert wird:

1. Dateninventur
    a. Welche Datenanwendungen werden im Unternehmen durchgeführt?
    b. Welche Daten existieren im Unternehmen?
2. Sensitivitätsanalyse
    a. Werden personenbezogene Daten gespeichert/gesammelt/verarbeitet?
    b. Existieren anderweitig sensible Daten im Unternehmen?
3. Aufsetzen eines geordneten Prozesses
    a. Wenn neue Datenanwendungen geplant werden.
    b. Wenn neue Datenquellen erschlossen werden, oder sich bestehende Datenlieferungen ändern
4. Reduktion
    a. Werden die Daten für die Datenanwendungen wirklich alle benötigt? Welche können weggelassen, bzw. gar nicht gesammelt werden?
    b. Umbau der Prozesse, falls sensible Daten gesammelt/verarbeitet werden, die nicht benötigt werden
5. Applikation von Schutzmechanismen (iterativer Prozess)
    a. Analyse, ob die Daten als Gesamtdatensatz anonymisiert, oder lediglich Auswertungsergebnisse benötigt werden.
    b. Wahl von Anonymisierungsparadigmen (bspw. k-anonymity) samt geeigneten Parametern.
    c. Qualitätskontrolle – Ist die Qualität der Datenauswertung mit anonymisierten Daten hoch genug?
    d. U.U. technische Optimierung des Anonymisierungsverfahrens – bspw. Outlier-Entfernung

    e. Aufbau einer sicheren Analyseumgebung mit striktem Logging, Trennung von Dataspaces, starkem Zugriffsrechtekonzept, Löschung von Daten nach der Verarbeitung
6. Löschen und Transparenz
    a. Erstellung eines Verarbeitungsmodells – Welche Daten fließen (auch aggregiert) in welche Ergebnisse ein?
    b. Applikation eines geeigneten Mechanismus um nachvollziehen zu können, welcher Datensatz wo einfloss – dies ermöglicht den direkten Zugriff für Auskunft und Löschung
    c. Wo technisch möglich: Schaffung von Prozess und technischen Tools zur physischen Löschung von Daten, bspw. Durch Überschreiben. Wo dies nicht möglich ist, Dokumentation wo und warum.

Aus derzeitiger Sicht dürfte die Umstellung auf die DS-GVO nur teilweise von Unternehmen wahrgenommen werden und auch wenn sie wahrgenommen wurde, ist der Umfang der Veränderungen nicht klar. Diese Wissensdefizite sollte die Politik durch gezielte Kampagnen möglichst schnell verändern. „Mit Hochdruck“ ist hier durchaus eine passende Formulierung.

Natürlich werden alle Organisationen die personenbezogene Daten verarbeiten, versuchen zu signalisieren, dass sie alle Gesetze einhalten. Für Außenstehende ist das schwer zu überprüfen. Hier könnte die Einführung eine schnellen Zertifizierung helfen in deren Rahmen ein Gütesiegel vergeben wird.

#### 6.3.2.2 Industriepolitisch

Bei den gegenwärtigen Marktverhältnissen gelingt es einigen Unternehmen eine dominante Stellung zu entwickeln und dadurch Marktmacht aufzubauen. Diese wiederum hilft diesen Unternehmen noch mehr Daten zu sammeln und damit ihre Marktposition zu festigen oder noch weiter auszubauen. Google, Facebook, Amazon sind die wichtigsten Beispiele für diese Entwicklung. Strikte Datenschutzbestimmungen könnten dazu führen, dass diese Unternehmen ihre Datenbestände nicht mehr im vollen Ausmaß auswerten könnten, wodurch es kleineren Mitbewerbern möglich sein sollte, konkurrenzfähige Angebote zu erstellen. Allerdings gibt es keine empirischen Belege, die diese Hypothese stützen.

Es ist wenig wahrscheinlich, dass die First Mover-Effekte, die diese Unternehmen generiert haben, durch die neue DS-GVO tatsächlich abgeschwächt werden. Wahrscheinlicher ist, dass die Unternehmen mit expliziter Zustimmung der NutzerInnen auf die Daten zugreifen können und strikte Datenschutzbestimmungen daher vor allem Marktneulinge stark treffen. Campbell et al. (2011) zeigen, dass gerade letzteres der Fall sein könnte, weil große, etablierte Unternehmen eher das Vertrauen von NutzerInnen erhalten, wenn es um die Einhaltung von Datenschutzstandards geht. Die strenge Regulierung von Kreditkarten in Neuseeland ist ein empirisches Beispiel dafür: die NachfragerInnen hatten den Eindruck, dass nur große etablierte Betreiber die Datenschutzrichtlinien einhalten konnten, und haben daher neue und kleine Anbieter gemieden.

Diese Argumentationslinie gilt vor allem für Produkte die sich an KonsumentInnen richten. Dort ist es tatsächlich schwer zu sehen, dass die etablierten Betreiber auch unter verschärften Datenschutzbestimmungen an Marktmacht verlieren und damit neue Betreiber eintreten können. Dennoch sollten auch in diesem Segment die NutzerInnen von verbessertem Daten profitieren, auch wenn die Marktmacht der quasi Monopolisten nicht wirklich eingedämmt wird.

Wenn die DS-GVO tatsächlich ein wirksames Instrument ist um ungewollte Datenweitergabe (auch für Werbezwecke) zu unterbinden – und damit auch das Datenschutz-Paradoxon

abschwächt – dann sollte es deutlich mehr Bezahlangebote und weniger „Daten für Produktnutzungs-Tauschgeschäfte“ geben.

### 6.3.2.3 Technologie- und Innovationspolitik

Die Hypothese, dass Datenschutzgesetze Innovationen bei Anbietern von Datenschutztechnologien stimulieren (wie z.B. Privacy Enhancing Technologies (PET= wie Verschlüsselung), bleibt aufrecht auch wenn in Europa derzeit die Rahmenbedingungen dafür nicht günstig sind, weil die Unsicherheiten und Zielkonflikte im Rahmen der DS-GVO groß sind und daher nicht klar ist, in welche Richtung sich die Technologien entwickeln können bzw. sollen. Diese Unsicherheit wird durch die Formulierungen und Forderungen der DS-GVO erzeugt und kann auch durch weitere Klarifikationen durch die entsprechenden Stellen korrigiert bzw. eingeschränkt werden.

Eine wesentliche Aufgabe für die öffentliche Hand ist die Beseitigung der Unsicherheiten, die die DS-GVO selbst mit sich bringt. Obwohl vieles ausjudiziert werden muss, gibt es auch andere Möglichkeiten mit den Unsicherheiten umzugehen. Dazu gehören weiterführende Erläuterungen oder die Option, verschiedene Ansätze mit den Behörden vorab zu diskutieren. Dazu braucht es aber die Bereitschaft der Behörden, die Dienstleistung zu erbringen und eine entsprechende Ressoucensausstattung. Wenn es gelingt die Unsicherheiten abzubauen, dann haben Entwickler von Datenschutztechnologien mehr Anreize in innovative Produkte und Dienstleistungen zu investieren, weil dadurch ein Markt geschaffen wird.

Dieser Ansatz ist konsistent in eine horizontale Strategie einzubauen, die die folgenden Bereiche umfasst:

- **Ausbildung:** ein Wettbewerbsvorteil durch besseren Datenschutz ist nur zu erzielen wenn die Nutzer dies auch wahrnehmen. Hierzu ist Awareness in der breiten Öffentlichkeit herzustellen und in der Ausbildung anzusetzen. Datenschutzaspekte sollten im Rahmen der Digitalisierungsstrategie „Schule 4.0“ eine wichtige Rolle spielen. Entsprechende Schulungen und Fortbildungsangebote sollten ebenfalls geschaffen werden.
- **Forschung:** In dieser Studie wird eine Reihe von Forschungsfragen vorgestellt, durch die eine effiziente Umsetzung der DS-GVO ermöglicht wird. Diese sollten in die Grundlagenforschung und angewandte Forschung Eingang finden (z.B. Informationserhaltung bei Anonymisierung, Ledgerarchitekturen zur Herstellung von Transparenz, Lösung des Widerspruchs Recht auf Vergessenwerden und Nachvollziehbarkeit). Förderung: Neben der entsprechenden Forschungsförderung z.B. durch die FFG sind auch Startups, die entsprechende Lösungen anbieten, zu fördern. Hier ist insbesondere eine Verbindung zu den Blockchain Aktivitäten (https://www.blockchain-austria.gv.at/, Blockchain Village) herzustellen, zumal durch diese Technologie eine von Grund auf andere Basis für die Verarbeitung personenbezogener Daten gelegt wird.
- **Gesetzliche Rahmenbedingungen:** Wie oben aufgezeigt ist eine adäquate Auslegung der Vorschriften für die Einwilligung entscheidend für die Effizienz von Big Data Anwendungen. Eine zu detaillierte Beschreibung zieht laufende Adaptionen der Einwilligungen und Unverständnis bei den Anwendern nach sich. Es sollten daher klar verständliche und nachvollziehbare Formulierungen für die Einwilligungserklärung gewählt werden. .

- Aufbau eines „MyData Local Hub“ in Österreich: Durch eine Mitgliedschaft beim MyData Projekt können mehrerer dieser Maßnahmen unterstützt werden:
    - Öffentliche Anbieter können ihre Expertise als Trusted Entitites im Rahmen eines Public Private Partnerships einbringen
    - Den Anwendern wird bewusst welche ihrer Daten wo gespeichert sind. Dadurch wird der verantwortungsvolle Umgang mit Einwilligungen gefördert, andererseits werden die Vertrauensprobleme kleinerer Anbieter verringert.
    - Die lokale Software Community kann durch den Open Source Charakter an der Entwicklung teilhaben und wird gefördert.
    - Die Erkenntnisse können auch für die Weiterentwicklung des eGovernment verwendet werden.

#### 6.3.2.3 Geopolitisch

Die DS-GVO ist gesellschaftspolitisch gewünscht, industriepolitisch interessant und als Teil einer geopolitischen Strategie unerlässlich. Europa hat digitale Technologien als gewöhnliche generische Technologien verkannt und die weiteren ökonomischen und geopolitischen Implikationen unterschätzt. Wohl aus der Verbundenheit gegenüber den USA, dem ständigen Drang wettbewerbsfähig zu bleiben und dem Versuch bestehende Rückstände aufzuholen, hat Europe vor allem auf die schnelle Diffusion von digitalen Technologien gesetzt und die Approprierung dieser vernachlässigt. Anstelle eines strategischen Ansatzes – wie etwa China oder Russland – wurde ein laissez faire Ansatz gewählt.

Gerade die Entwicklungen in den letzten Monaten haben gezeigt, dass diese Positionierung nicht tragfähig ist und es eine eigenständige und europazentrierte Herangehensweise braucht. Europa ist in vielen der heißen Themen rund um die Digitalisierung nicht vertreten (z.B. AI im militärischen Bereich). Diese Vakui kann man sich nicht auf Dauer leisten, weil man ohne diese „hand on“ Expertise auch nicht bei der Gestaltung der Rahmenbedingungen mitarbeiten kann. Damit man hier Fortschritte macht, muss die europäische Dimension bei diesen Fragen deutlich stärker ausgebaut werden. Das ist zwar schwierig, aber die Rahmenbedingungen sind mit einer funktionierenden deutsch-französischen Achse deutlich besser als noch vor Kurzem. Die österreichische Positionierung über die tagesaktuellen Themen hinaus ist hier nicht zu erkennen. Es wäre wünschenswert hier die europäische Dimension aktiv zu stärken und so an der geopolitischen Neupositionierung mitzuarbeiten.

Klar ist, dass die derzeitigen Schritte auf europäischer Ebene – Verfolgung von Marktmachtmissbrauch, Einhaltung der Datenschutzbestimmungen – Defensivstrategien sind und als solche zwar wichtig, aber nur dann erfolgreich sein werden, wenn auch aktive Elemente eingebaut werden. Die DS-GVO kann ein Aktivposten werden, wenn sie auch durch andere Politikmaßnahmen im Sinne einer horizontalen Strategie gestützt wird und die NutzerInnen wieder Verfügungsgewalt über ihre Daten haben. Am besten kann diese durch System erfolgen, wo NutzerInnen den Zugriff auf Ihre Daten zentral verwalten können und alle Abfragen dokumentiert werden. Dazu sind einheitliche Standards und APIs für Europa zu erarbeiten und möglichst schnell zu etablieren.

Insgesamt muss sich aber auch Europa als Modell für wirksamen Datenschutz positionieren, wenn man die ökonomischen Chancen realisieren will. Dies heißt auch, dass man Personen die das europäische Datenschutzniveau „genießen“ wollen eine Art eResidency gewährt wie es bereit Estland macht. Österreich könnte hier nachziehen und eine flächendeckende europäische Lösung vorwegnehmen. Letztendlich müssen – wie bereits erwähnt – alle Unternehmen die DS-

GVO einhalten, wenn sie europäische KundInnen haben. Wenn man sie Zahl der datenschutzrechtlichen Europäer über eResidency erhöht, dann verbreitet sich europäisches Datenschutzrecht über Europa hinaus und kann so zum Standard werden. Man kann auch überlegen, mit Länder zwischen den geopolitischen Machtblöcken (z.B. Südamerika, Japan) Allianzen zur Übernahme der europäischen Datenschutzregeln zu bilden[35].

Abschließend – und auf Innovation bezogen - wurde die DS-GVO nicht konzipiert, um Innovation zu verhindern, sondern um Datenschutz zu verbessern. Das ist ebenfalls eine Innovation, die mit beachtlichen Kosten für die Organisationen, die sie implementieren müssen, verbunden ist. Daher sollte man jetzt alle begleitenden Maßnahmen setzen, damit die Umstellung möglichst einfach über die Bühne geht und erfolgreich verläuft. Dazu gehören Informationskampagnen für die Organisationen die dazu verpflichtet sind als auch für NutzerInnen. Aufklärung ist auch bei den KonsumentInnen notwendig, damit diese die neuen Möglichkeiten annehmen und damit umgehen können. Nur wenn informierte NutzerInnen von den neuen Möglichkeiten Gebrauch machen, kann das oft beachtete Daten-Paradoxon verhindert werden.

Die DS-GVO bringt Einschränkungen für Big Data Strategien, wenn personenbezogene Daten betroffen sind. Dies gilt nicht wenn Unternehmensdaten oder Nutzungsdaten verarbeitet werden. Klar ist auch, dass bei Innovationen und Big Data Analysen nicht taxativ aufgezählt werden kann, welche Daten notwendig sind. Hier liegt der Ball eindeutig bei den Organisationen, die personenbezogene Daten verarbeiten wollen. Nur wenn sie aus der Sicht der „Datenlieferanten“ eine vertrauenswürdige Organisation sind, werden breit formulierte Big Data- und Innovationsstrategien über entsprechend gefasste Einverständniserklärungen auf Akzeptanz stoßen.

Europa ist aber auch gefordert, die strengen Datenschutzbestimmungen auch in Abkommen mit Drittstaaten einzuhalten. Dies gilt natürlich für die DS-GVO selbst, aber auch für das aufgehobene Safe Harbor Abkommen mit den USA und anderen Datenaustauschübereinkommen (z.B. Fluggastdaten) bei den die bisherigen Bestimmungen deutlich zu weit gingen. Europa muss hier auf der ganzen Linie konsequent sein, die eigenen Gesetze auf allen Ebenen einhalten und damit seine Autonomie wahren.

[35] Das wäre deutlich sinnvoller als – wie derzeit öfters angedacht – den Datenaustausch über bi- und multinationale Handelsabkommen zu ermöglichen.

# 7 Verwendete Literatur


Accorsi, R., On the relationship of privacy and secure remote logging in dynamic systems. In IFIP International Information Security Conference, 2006

Acquisti, A., Adjerid, I., Brandimarte, L., "Gone in 15 seconds: The limits of privacy transparency and control." IEEE Security & Privacy 11.4 (2013): 72-74.

Acquisti, A., Taylor, C., Wagman, L., the Economics of Privacy, Journal of Economic Literature, Vol. 52, No. 2, 2016.

Acquisti, A., H. R. Varian (2005). Conditioning prices on purchase history. Marketing Science 24 (3), 367{381.

Adams, T., Surge Pricing Comes to the Supermarket, Guardian, 4. Junie 2017, https://www.theguardian.com/technology/2017/jun/04/surge-pricing-comes-to-the-supermarket-dynamic-personal-data.

Albrecht, J. P., & Jotzo, F. (2017). Das neue Datenschutzrecht der EU. Grundlagen, Gesetzgebungsverfahren, Synopse, Baden-Baden.

Arieli, D., Predictably Irrational, The Hidden Forces That Shape Our Decisions, Harper Collins Publisher, 2010.

Article 29 Data Protection Working Party (2004), Opinion 10/2004 on More Harmonised Information Provisions: Adopted on 25th November 2004. 11987/04/EN. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp100_en.pdf.

Bartolini, C., Muthuri, R., Cristiana, S., "Using ontologies to model data protection requirements in workflows", 2015.

Bellare M. and B. Yee. Forward integrity for secure audit logs. Technical report, Technical report, Computer Science and Engineering Department, University of California at San Diego, 1997.

Bernd, M., Kieseberg, P., Weippl, E. R., Holzinger, A., "The Right to Be Forgotten: Towards Machine Learning on Perturbed Knowledge Bases," in International Conference on Availability, Reliability, and Security, 2016.

Blank, S., Dorf, B., The Startup Owner´s Manual, K&S Ranch Press, 2012.

Blättel-Mink B., Menez R., Kompendium der Innovationsforschung, 2015.

Bonatti, P. A., & Olmedilla, D., Rule-based policy representation and reasoning for the semantic web. Proceedings of the Third International Summer School Conference on Reasoning Web, RW'07, pp. 240-268. Springer-Verlag, Berlin, Heidelberg, 2007.

Bonatti P., De Capitani di Vimercati, S., Samarati, P., An algebra for composing access control policies. ACM Transactions on Information and System Security (TISSEC) , 5(1), 2002.

Bonatti P., Kirrane S., Polleres A., and Wenning R. Transparent personal data processing: The road ahead. In TELERISE: 3rd International Workshop on TEchnical and LEgal aspects

of data pRIvacy and SEcurity @ SAFECOMP2017, Trento, Italy, September 2017. to appear

Bradshaw, J. M., Dutfield, S., Benoit, P., & Woolley, J. D. (1997). Software agents. MIT Press, Cambridge, MA, USA, Ch. KAoS: Toward an Industrial-strength Open Agent Architecture, pp. 375–418.

Cadwalladr, C., The great British Brexit robbery: how our democracy was hijacked, Guardian, 7. 5. 2017, https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy

Campbell, J. D., Goldfarb, A., Tucker, C., Privacy Regulation and Market Structure. mimeo, University of Toronto, 2011.

Casadesus-Masanell, Ramon, and Andres Hervas-Drane, "Competing with privacy." Management Science 61.1 (2015): 229-246.

Chesbrough, H., Open Innovation, The New Imperative for Creating and Profiting from Technology. 2003

Chesbrough H., Brunswicker S., Managing Open Innovation in Large Firms, Survey Report, Executive Survey on Open Innovation, Fraunhofer 2013.

Croll, A., Yoskovitz, B., Lean Analytics, Use Data to Build a Better Startup Faster, O´Reilly, Sebastopol, 2013.

Determann, L., Adequacy of data protection in the USA: myths and facts. International Data Privacy Law 2016; 6 (3): 244-250, 2016.

Dwork, C., "Differential privacy: A survey of results." In International Conference on Theory and Applications of Models of Computation, pp. 1-19. Springer Berlin Heidelberg, 2008.

Evans, P. C., Gawner , A, The Rise of the Platform Enterprise, A Global Survey, The Center for Global Enterprise, 2015, http://thecge.net/wp-content/uploads/2016/01/PDF-WEB-Platform-Survey_01_12.pdf

Fassauer, R., "Personalisierung im E-Commerce–zur Wirkung von E-Mail-Personalisierung auf ausgewählte ökonomische Kennzahlen des Konsumentenverhaltens." (2014).

Felt, A.P., Ha, E., Egelman, S., Haney, A., Chin, E. & Wagner, D., Android Permissions: User Attention, Comprehension, and Behavior. Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS 2012. ACM, 2012.

Fernández J., Kiesling, E., Kirrane, S., Neuschmid, J., Mizerski, M., Polleres, A., Sabou, M., Thurner, T., Wetz, P., Propelling the Potential of Enterprise Linked Data in Austria: Roadmap and Report. edition mono/monochrom, Zentagasse 31/8, A-1050 Vienna, Austria, December 2016.

Fernández J., S. Kirrane, A. Polleres, and S. Steyskal, Self-enforcing access control for encrypted RDF. In Proceedings of the 14th European Semantic Web Conference (ESWC2017) , Portorož, Slovenia, May 2017. URL http: //polleres.net/publications/fern-etal-ESWC2017.pdf .

Fernández Garcia, J. D., Umbrich, J., Knuth, M. Polleres, A., Evaluating query and storage strategies for RDF archives. In 12th International Conference on Semantic Systems (SEMANTICS) , ACM International Conference Proceedings Series, 2016.

Fruehwirt, P., Huber, M., Schmiedecker, M., Weippl, E. R., "InnoDB Database Forensics," in Proceedings of the 24th International Conference on Advanced Information Networking and Applications, 2010.

Fruehwirt, P., Kieseberg, P., Schrittwieser, S., Huber, M., Weippl, E.R., "InnoDB Database Forensics: Reconstructing Data Manipulation Queries from Redo Logs," in The Fifth International Workshop on Digital Forensics (WSDF), 2012.

Fruehwirt, P., Kieseberg, P., Schrittwieser, S., Huber, M., Weippl, E. R., "InnoDB Database Forensics: Enhanced Reconstruction of Data Manipulation Queries from Redo Logs," Information Security Technical Report (ISTR), Special Issue: ARES, 2013.

Fruehwirt, P., Kieseberg, P., Krombholz, K, Weippl, E. R., "Towards a forensic-aware database solution: Using a secured database replication protocol and transaction management for digital investigations," Digital Investigation, vol. 11, pp. 336-348, 2014.

Fruehwirt, P., Kieseberg, P., Weippl, E. R., "USING INTERNAL MySQL/InnoDB B-TREE INDEX NAVIGATION FOR DATA HIDING." In IFIP International Conference on Digital Forensics, pp. 179-194. Springer International Publishing, 2015.

Goldfarb, A. and C. Tucker (2011a). Online display advertising: Targeting and obtrusiveness. Marketing Science 30 (3), 389{404.

Goldfarb, A. and C. Tucker (2011b). Privacy regulation and online advertising. Management Science 57 (1), 57{71.

Gottlieb, D. and K. Smetters (2011). Grade non-disclosure. Available at NBER: http://www.nber.org/papers/w17465.

Gross-Amblard, D. (2003). Query-preserving watermarking of relational databases and XML documents. SIGART Symposium on Principles of Database Systems

Gürses, S., del Alamo, J. M., "Privacy Engineering: Shaping an Emerging Field of Research and Practice." IEEE Security & Privacy 14.2 (2016): 40-46.

Cadwalladr, C., The great British Brexit robbery: how our democracy was hijacked, Guardian, 7. 5. 2017, https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy

Hansen, M., "Data Protection by Design and by Default à la European General Data Protection Regulation." Privacy and Identity Management. Facing up to Next Steps. Springer International Publishing, 2016. 27-38.

Hayek, F., The Use of Knowledge in Society, American Economic Review, 1945, https://assets.aeaweb.org/assets/production/journals/aer/top20/35.4.519-530.pdf

Hedbom, H., Pulls, T., Hjärtquist, P. & Lavén, A. (2009). Adding secure transparency logging to the prime core. Privacy and Identity Management for Life, pp. 299-314. Springer Berlin Heidelberg

Hippel E., The Sources of Innovation. 1988

Holt J. E. Holt. Logcrypt: forward security and public verification for secure audit logs. In Proceedings of the 2006 Australasian workshops on Grid computing and e-research-Volume 54 , 2006.. Logcrypt: forward security and public verification for secure audit logs. In Proceedings of the 2006 Australasian workshops on Grid computing and e-research-Volume 54 , 2006

Holtz, L. E., Zwingelberg, H., & Hansen, M. (2011) Privacy Policy Icons. Privacy and Identity Management for Life, pp. 279-285

Hope-Bailie, A., Thomas, S., Interledger: Creating a standard for payments. In Proceedings of the 25th International Conference Companion on World Wide Web , 2016.

Huizingh E., Open innovation: State of the art and future perspectives. In: Technovation 31, 2011, 2-9.

Jahnel, D. (2010). Handbuch Datenschutzrecht. Auflage, Salzburg.

Kagal, L., & Finin, T. (2003). A policy language for a pervasive computing environment. Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks, pp. 63-74. IEEE Comput. Soc.

Kelley, P. G., Cesca, L., Bresee, J., Cranor, L. F., Standardizing privacy notices: An online study of the nutrition label approach. Mimeo, Carnegie Mellon University CyLab CMU-CyLab-09-014, 2011.

Kieseberg, P., Schrittwieser, S., Schmiedecker, M. Huber, M., Weippl, E. R., "Trees Cannot Lie: Using Data Structures for Forensics Purposes," in European Intelligence and Security Informatics Conference (EISIC 2011), 2011.

Kieseberg, P., Schrittwieser, S., Mulazzani, M., Echizen, I., Weippl, E. R., "An algorithm for collusion-resistant anonymization and fingerprinting of sensitive microdata." Electronic Markets 24, no. 2 (2014): 113-124

Kieseberg, P., Malle, B., Fruehwirt, P., Weippl, E. R., Holzinger, A., "A tamper-proof audit and control system for the doctor in the loop," Brain Informatics, pp. 1-11, 2016

Kokolakis, S., "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon." Computers & Security, 2015.

Kübler, D., Privates öffentlich: Datenschutz ist dem Schnäppchenjäger nichts wert. In: WZB-Mitteilungen, 132, pp. 10-11, 2011 URN: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-3087

Kirrane, S., Mileo, A., Decker, S., Access control and the resource description framework: A survey. Semantic Web , 8(2):311–352, 2017.

Langelaar, G., Setyawan, I., & Lagendijk, R. (2000). Watermarking digital image and video data. A state-of-the-art overview. IEEE Signal Processing Magazine, 17(5), 20–46

Leo, H. Digitalisation and innovation – how new technologies can help to overcome economic barriers, paper presented at OSCE conference “Towards the Vision of a Common Economic Space from Vancouver to Vladivostok: Connectivity, Trade and Economic Cooperation”, 15-16 May 2017, Linz, Austria.

Leo, H., Seethaler, U., Schritte zur Operationalisierung der Open Innovation Strategie für Österreich, Wien, 2017.

Leo, H., Palme, G., Volk, E., Die Innovationstätigkeit der österreichischen Industrie, Technologie- und Innovationstest 1990, Wifo, Wien. 1992.

Li, W., Yuan, Y., Li, X., Xue, X., & Lu, P. (2005). Overview of digital audio watermarking. Tongxin Xuebao (Journal on Communications), 26(2)

Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "t-closeness: Privacy beyond k-anonymity and l-diversity." In 2007 IEEE 23rd International Conference on Data Engineering, pp. 106-115. IEEE, 2007.

Liu, Siyuan, Shuhong Wang, Robert H. Deng, and Weizhong Shao. "A block oriented fingerprinting scheme in relational database." In International Conference on Information Security and Cryptology, (2004): 455-466

Madden, M. Rainie, L., (2015). Americans' attitudes about privacy, security and surveillance.

Martin, K D., Murphy, P. E., "The role of data privacy in marketing." Journal of the Academy of Marketing Science (2016): 1-21.

Ma D., Tsudik, G., A new approach to secure logging. ACM Transactions on Storage (TOS) , 5(1), 2009.

Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M., "l-diversity: Privacy beyond k-anonymity." ACM Transactions on Knowledge Discovery from Data (TKDD) 1, no. 1 (2007): 3

Maurya, A., Running Lean: Iterate from Plan A to a Plan That Works, O'Reilly, 2012.

Mazzucato, M., The Entrepreneurial State: debunking public vs. private sector myths, Anthem Press, London, 2013.

McDonald, A. M., Cranor, L. F., "The cost of reading privacy policies." ISJLP 4 (2008): 543.

McRoberts, M., Doncel, R.V., ODRL Version 2.1 Ontology, 2015 Available at: http://www.w3.org/ns/odrl/2/ODRL21.

Newman, D., The Process of Design Squiggle, Central Office of Design.

Noble, Ch. H., Durmusoglu, S. S., Griffin, A., Open Innovation, New Product Development essentials from the PDMA, Wiley, Hoboken, 2014.

Oetzel, M. C., Spiekermann, C., "Privacy-by-Design through Systematic Privacy Impact Assessment-a Design Science Approach." ECIS. 2012.

Paal, B. P., Pauly, D. A., & Ernst, S. (2017). Datenschutz-Grundverordnung.

Peeters R., T. Pulls, and K. Wouters. Enhancing transparency with distributed privacy-preserving logging. In ISSE 2013 Securing Electronic Business Processes. Springer, 2013.

Piller F., Lüttgens D., Pollok P., Open Innovation. Methoden und Erfolgsbeurteilung, in: Wirtschaftswissenschaftliches Studium Zeitschrift für Ausbildung und Hochschulkontakt, 11-2013.

Piller F., Not Invented here. In: Interview von Ramge Thomas, Brand Eins 01-2017, S. 72.

Pulls T., R. Peeters, and K. Wouters. Distributed privacy-preserving transparency logging. In Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society , 2013.

Rainie, L., S. Kiesler, R. Kang, M. Madden, M. Duggan, S. Brown, and L. Dabbish (2013). Anonymity, privacy, and security online. Pew Research Center.

Ries, E., The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses, 2011.

Rinne M., E. Blomqvist, R. Keskisärkkä, and E. Nuutila. Event processing in rdf. In Proceedings of the 4th International Conference on Ontology and Semantic Web Patterns-Volume 1188 , 2013.

Rogaway, P., "A synopsis of format-preserving encryption." In UNPUBLISHED MANUSCRIPT. 2010.

Sackmann S., J. Strüker, and R. Accorsi. Personalization in privacy-aware highly dynamic systems. Communications of the ACM , 49(9), 2006.

Samuel, J., Zhang, B., RequestPolicy: Increasing web browsing privacy through control of cross-site requests. Privacy enhancing technologies. Springer Berlin Heidelberg, 2009.

Schneier B., Kelsey, J., Cryptographic support for secure logs on untrusted machines. In USENIX Security , 1998.

Schneier B., Kelsey, J., "Secure audit logs to support computer forensics." ACM Transactions on Information and System Security (TISSEC) 2, no. 2 (1999): 159-176.

Schumpeter, J. A., Capitalism, Socialism, and Democracy, Harper, New York, 1942.

Scott, M., Europe's tech ambition: To be the world's digital policeman, The Continent's policy-makers want to determine how companies and their users behave online, Politico, 8/20/17, http://www.politico.eu/article/europe-tech-ambition-to-be-world-digital-police-man/

Scott, M., Hirst, N., Europe´s next competition clash: Online data, Politico.eu, 8/25/17, http://www.politico.eu/article/europe-competition-google-amazon-facebook-data-privacy-antitrust-vestager/

Schwartz, P. M., Preemption and privacy. Yale Lj, 118, 902, 2008.

Seneviratne, O., Kagal, L., Enabling privacy through transparency. In Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on Privacy, Security and Trust, 2014.

Seth, G., Lynch, N., “Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services.” ACM SIGACT News, v. 33 issue 2, 2002, p. 51–59.

Shellshear, E., Innovation Tools, The most successful tools to innovate effectively and cheaply, 2016.

Singla, S., Kumar, R., Kumar, D., Natural Computing for Automatic Test Data Generation Approach Using Spanning Tree Concepts. Procedia Computer Science, 85, 929-939, 2016

Sion, R., Atallah, M., Prabhakar, S., Watermarking relational databases, 2002.

Skopik, F., Settanni, G., Fiedler, R., Friedberg, I., Semi-synthetic data set generation for security software evaluation. In 2014 Twelfth Annual International Conference on Privacy, Security and Trust, Toronto, ON, Canada, July 23-24, 2014, pages 156{163, 2014

Spiekermann, S., "The challenges of privacy by design." Communications of the ACM 55.7 (2012): 38-40.

Som O., Jäger A., Maloca S., Open Innovation - ein universelles Erfolgskonzept? in: Modernisierung der Produktion. Mitteilung aus der ISI-Erhebung. Frauenhofer, Ausgabe 66, 08-2014.

Steyskal, S., Polleres, A., Towards formal semantics for ODRL policies. In 9th International Web Rule Symposium (RuleML2015) , number 9202, pages 360–375, Berlin, Germany, Aug. 2015. Springer. URL http://www.polleres.net/publications/stey-poll-2015RuleML.pdf .

Sweeney, L., "k-anonymity: A model for protecting privacy." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10, no. 05 (2002): 557-570.

Sweeney, L., "Comments to the department of health and human services on “standards of privacy of individually identifiable health information”.", 2002.

Taylor, C. R., Consumer privacy and the market for customer information. RAND Journal of Economics 35 (4), 631{650, 2004.

Tikkinen-Piri, C., Rohunen, A., Markkula, J., EU General Data Protection Regulation: Changes and implications for personal data collecting companies. Computer Law & Security Review, 2017.

Tucker, C. E., The economics of advertising and privacy. International journal of Industrial organization 30 (3), 326{329, 2012.

Turow, J., J. King, C. Hoofnagle, A. Bleakley, and M. Hennessy (2009). Americans reject tailored advertising and three activities that enable it. Working paper.

Vayena, E., Mastroianni, A., Kahn, J., 2013. Caught in the web: informed consent for online health research. Sci Transl Med, 5(173), p.173fs6.

Licenses Compatibility and Composition in the Web of Data. Proceedings of Third International Workshop on Consuming Linked Data, COLD, 2012.

Wagner P., Piller T., Open Innovation - Methoden und Umsetzungsbedingungen, in: Howald J. (Hg.) et.al., Innovationsmanagement 2.0, 2011.

Weitzner D. J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., Sussman, G. J., Information accountability. Communications of the ACM , 51(6), 2008.

Wolff, H. A., & Brink, S., Beck'scher Online-Kommentar Datenschutzrecht. Aufl. München: CH Beck, 2013.

Wouters K., K. Simoens, D. Lathouwers, and B. Preneel. Secure and privacy-friendly logging for egovernment services. In Availability, Reliability and Security, 2008. ARES 08. Third International Conference on Availability, Reliability and Security, 2008.

Xiao, X., Tao, Y., "M-invariance: towards privacy preserving re-publication of dynamic datasets." In Proceedings of the 2007 ACM SIGMOD international conference on Management of data, pp. 689-700. ACM, 2007

Yang, X-C., Liu, X-Y, Wang, B., Yu, G., "K-anonymization approaches for supporting multiple constraints." Ruan Jian Xue Bao(Journal of Software) 17, no. 5 (2006): 1222-1231.

Zarsky, T. Z., The Privacy-Innovation Conundrum, Lewis & Clark Law Review, 2015.

Zuckerberg, M., Building Global Community, 2017, https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/

Zuiderveen Borgeswius, F. J., Security & Privacy, 'Informed Consent. We Can Do Better to Defend Privacy', IEEE (Volume 13, Issue 2, p. 103-107).